Mixture Models in Financial Risk Modeling

Dissertation
submitted to the Faculty of Economics,
Business Administration and Information Technology
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Lars Jochen Krause
from Germany

approved in February 2014 at the request of
Prof. Dr. Marc S. Paolella
Prof. Dr. Erich W. Farkas

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zürich, 12.02.2014

Chairman of the Doctoral Board: Prof. Dr. Josef Zweimüller

To my family.

Contents

I Introduction 1

II Research Papers 9

1 Stable Mixture GARCH 11
- 1.1 Introduction . . . 13
- 1.2 Stable Mixture GARCH . . . 16
- 1.3 Estimation of Mixture GARCH Models . . . 24
- 1.4 Univariate Empirical Results . . . 28
- 1.5 ICA-MixStable-GARCH . . . 33
- 1.6 Conclusions and Future Research . . . 36
- Appendix . . . 37

2 Time-varying Mixture GARCH 55
- 2.1 Introduction . . . 57
- 2.2 Mixed Normal GARCH . . . 59
- 2.3 Time-varying weights . . . 60
- 2.4 Empirical Results . . . 69
- 2.5 Conclusions and Further Extensions . . . 72
- Appendix . . . 73

3 Portfolio optimization using the noncentral t 85
- 3.1 Introduction . . . 87
- 3.2 Saddlepoint approximation of expected shortfall for transformed means . . . 88
- 3.3 Portfolio optimization using the noncentral t . . . 90
- 3.4 Conclusion . . . 97
- Appendix . . . 98

III Curriculum Vitae 113

List of Tables

1.1 Small sample performance analysis of the studied estimators 40
1.2 Profile log likelihood study of the GARCH power parameter 41
1.3 Test statistics on the optimal choice of the GARCH power parameter 42
1.4 Estimated parameter values and standard errors 43
1.5 Estimated log likelihood, AIC and BIC values 44
1.6 Out-of-sample forecast performance results I 45
1.7 Out-of-sample forecast performance results II 46
1.8 Predicted VaR coverage percentages 47
1.9 Integrated root mean squared errors 48
1.10 Christoffersen test statistics 49

2.1 Estimated log likelihood values 76
2.2 Estimated BIC values 76
2.3 Out-of-sample performance comparison results 77
2.4 Predicted VaR coverage percentages 78
2.5 Integrated root mean squared errors 78
2.6 Christoffersen test statistics 79
2.7 Comparison of two variants of the new model 80

3.1 Annualized portfolio performance measures 103
3.2 Average realized predictive log likelihood values 103
3.3 Out-of-sample forecast performance results for the min-ES portfolio 104
3.4 Out-of-sample forecast performance results for the equally-weighted portfolio 104

List of Figures

1.1 The Stable Paretian and Nolan's density approximation 22
1.2 The impact of mixture degeneracy for the estimators under study 39

2.1 The news impact curve under different models 74
2.2 Estimated news impact curves 75

3.1 Average computation times for the min-ES portfolio 93
3.2 Illustration of the behavior of terms included in the infinite sum in (3.17) 101
3.3 Annual performance results for the min-ES portfolio 105
3.4 Cumulative returns and predictions of mean and variance for the min-ES portfolio 106
3.5 Predictions of median, value-at-risk and expected shortfall for the min-ES portfolio . . . 107
3.6 Evolution of the degrees of freedom parameter over time 108
3.7 Evolution of the noncentrality parameters over time 108
3.8 Predictions of mean and variance for the equally-weighted portfolio 109
3.9 Predictions of median, value-at-risk and expected shortfall for the 1/K portfolio 110

Part I

Introduction

Acknowledgments

My thanks go to my family, friends and colleagues who have contributed to this thesis in numerous ways throughout the years. Especially to Marc Paolella, Stefan Mittnik, Markus Haas, Simon Broda, Sven Steude, Pawel Polak, Kerstin Kehrle, Tatjana Puhan, Ingo Goetze and, last but not least, Katharina Schönau. Of course, the biggest thank-you goes to my doctoral adviser Prof. Marc Paolella! Marc, this work would not have been possible without your guidance, patience and continued support. My special thanks! Finally, I also wish to express my gratitude to the Department of Banking and Finance of the University of Zurich (highlighting our IT team), as well as to the Swiss National Science Foundation!

Introduction

Risk management is a crucial ingredient of sustainable development in a world that is becoming increasingly quantitative while being continuously shaken by financial bubbles and crashes of varying magnitude. The recent and enduring hype on *big data*, including the colossal amounts of high-frequency financial returns data available today, stresses a trend towards a new form of high-level technical (risk) analysis that is not only present in (empirical) finance but basically in every discipline in academia and industry. Concerning finance and the financial sector at large, it is no overstatement to say that an elaborated quantitative risk management incorporating such data could have prevented, or at least attenuated, some of the recent financial crises. However, for a profound risk management, adequate measures of risk are indispensable comprising models, methods and techniques that characterize and capture the particular type of risk involved in a certain business activity; e.g., the risk to default, or the risk of loosing a certain amount of an investment for a given probability of occurrence (downside risk). The latter, the downside risk, is a cornerstone of asset allocation, and as such, of high value for virtually all financial institutions such as hedge funds, pension funds, or insurance companies where risk evaluation and management is a critical element in business. For the insurance sector, for example, according to CEIOPS (2010) and EIOPA (2011), one-quarter of the overall risk belongs to financial equity risk whose accurate assessment is one of the key aspects of this thesis. On the other hand, regulatory agencies specify procedures and parameters for companies for determining the downside risk in terms of capital requirements to, among other targets, reduce the likelihood of firms defaulting; see the regulatory frameworks Basel III and Solvency II of the European Union. Roughly summarized, quantifying the downsize risk has evolved into daily practice and has inevitably become one of the most important factors in determining the level of resources expended in order to mitigate risk.

Stressing that the best risk measure is worth nothing if the underlying statistical model and the employed estimation methods and numerical techniques are not appropriate, this thesis contributes to the improvement of risk measurement by introducing new models, methods and techniques dedicated to predict the density of tomorrow's return more accurately. The goal is to produce accurate return density forecasts for financial assets from which the distribution and, thus, risk measures can be easily derived. Special emphasis is thereby set on (computational) feasibility, accuracy and numerical reliability throughout the chapters. Considering modeling the predictive density, i.e., the evolution of asset returns over time, the class of mixture generalized autoregressive conditional heteroscedasticity (mixture GARCH) models is chosen; either based on finite mixture distributions (see Haas et al., 2004; and Alexander and Lazar, 2006; and also Haas et al., 2009), or location scale mixtures of normals (see, e.g., Mencía and Sentana, 2009; and Jondeau, 2010). Mixture distributions possess a successful and long history in the literature (for finite mixtures see, e.g., Pearson, 1894) and have been shown to be well suited for capturing stylized facts of financial asset returns, which refer to empirically observed phenomena common to all such data sets. Finite mixtures, in particular, being convex combinations of density functions, can flexibly mimic all kinds of functional shapes and therefore lend themselves to account for stylized facts such as non-normality, heavy-tails, asymmetry, and non-ellipticity; while GARCH filter (see Bollerslev, 1986; and for ARCH Engle, 1982) are well-known to excellently capture time-varying volatilities, volatility clusters and volatility persistence. For judging about the quality of the approaches devised, extensive out-of-sample forecast comparisons with respect to one-day ahead density and risk forecasts are conducted which demonstrate the applicability and usefulness on historic market data. In a nutshell, the proposed models, methods and techniques are shown to deliver out-of-sample, future forecasts that outperform all competitors in our comparisons.

Chapter 1 comprises the working paper Broda et al. (2011) that complements the published version Broda et al. (2013) by adding computational details (like an improved method for evaluating the stable Paretian density) and extended discussions. The paper contributes three-fold. (i) Building on Haas et al. (2004) it introduces a new class of mixture GARCH models for univariate returns using finite mixtures of stable Paretian distributions. The class nests numerous models currently in use which are shown to be outperformed by the general model. (ii) Nouveau maximum likelihood estimators are devised that solve the mixture degeneracy problem of the standard maximum likelihood estimator in both the unconditional (no GARCH) as well as the conditional case where the latter allows for mixture components with GARCH effects. The presented solution also carries over to multivariate settings. (iii) It outlines an independent component analysis framework for use with univariate (mixture) models that, given the tractability of the relevant characteristic functions, facilitates portfolio optimization by minimizing the expected shortfall of the (predicted) portfolio return where the portfolio return is modeled as a weighted sum of mixture distributions.

Chapter 2 shows the working version, Haas et al. (2013a), of the published paper Haas et al. (2013b) which is a highly altered and extended revision of the manuscript Haas et al. (2006). The paper builds on Broda et al. (2013) and contributes two-fold. (i) It proposes a new mixture GARCH variant with time-varying mixing weights that facilitates an empirically suitable representation of Engle and Ng's (1993) news impact curve with an asymmetric (i.e., negatively correlated) impact of the unexpected return shock on future volatility, commonly referred to in the literature as Black's leverage effect. Among the various models studied, the best performing one relates mixing weights at time t to past returns and past realized likelihood values at time $t-1$, thus, suggesting that the leverage effect in financial returns data is closely connected to the time-varying interplay of different groups of market participants represented by the mixture components. An out-of-sample comparison confirms the superiority of the new model over asymmetric GARCH models such as E-GARCH and GJR-GARCH, to name just two. (ii) It gives an algorithm for the fast computation of optimal mixing weights which plays a crucial role in the construction of the best performing model. This *reduced* EM algorithm is general and applicable in the estimation of unconditional mixtures as well as in combination with the estimators in Broda et al. (2013).

Chapter 3 corresponds to a working paper building on Broda and Paolella (2010). The paper contributes two-fold. (i) It introduces a new multivariate GARCH model based on the multivariate noncentral t distribution and Engle's DCC filter. The model accounts for most stylized facts of asset returns, including time-varying volatility and correlation, fat tails and asymmetry, as well as non-ellipticity, and features a less extreme tail dependence behaviour compared to the multivariate generalized hyperbolic, see, e.g., Jondeau (2010). Using the main result of Broda and Paolella (2010), a closed form expected shortfall (ES) expression for the distribution of portfolio returns for use with the result in Rockafellar and Uryasev (2000) is obtained that significantly reduces computation times in the minimum ES portfolio optimization. (ii) For the estimation of the multivariate noncentral t distribution, a new density approximation is devised based on which a three-step estimation procedure for the portfolio model is developed. Computational details are worked out and the proposed model is shown to outperform the classic DCC model in terms of out-of-sample density forecast quality.

Bibliography

Alexander, C. and Lazar, E. (2006). Normal mixture GARCH(1, 1): applications to exchange rate modelling. *Journal of Applied Econometrics*, 21(3):307–336.

Black, F. (1976). Studies in Stock Price Volatility Changes. In *American Statistical Association, Proceedings of the Business and Economic Statistics Section*, pages 177–181.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2011). Stable mixture GARCH Models. *Swiss Finance Institute (SFI) Research Paper No. 11-39*.

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013). Stable mixture GARCH models. *Journal of Econometrics*, 172(2):292–306.

Broda, S. A. and Paolella, M. S. (2010). Saddlepoint Approximation of Expected Shortfall for Transformed Means. *UvA Econometrics Discussion Paper 2010/08. University of Amsterdam*.

CEIOPS (2010). QIS5 Calibration Paper. CEIOPS-SEC-40-10, Committee of the European Insurance and Occupational Pension Supervisors.

EIOPA (2011). EIOPA report on the fifth Quantitative Impact Study for Solvency II. EIOPA-TFQIS5-11/001, European Insurance and Occupational Pensions Authority.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica,*, 50(4):987–1007.

Engle, R. F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics*, 20:339–350.

Engle, R. F. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48(5):1749–1778.

Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013a). Time-varying Mixture GARCH Models and Asymmetric Volatility. *Swiss Finance Institute (SFI) Research Paper No. 13–04*.

Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013b). Time-varying Mixture GARCH Models and Asymmetric Volatility. *North American Journal of Economics and Finance*, 26:602–623.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.

Haas, M., Mittnik, S., and Paolella, M. S. (2009). Asymmetric Multivariate Normal Mixture GARCH. *Computational Statistics & Data Analysis*, 53(6):2129–2154.

Haas, M., Mittnik, S., Paolella, M. S., and Steude, S. C. (2006). Analyzing and Exploiting Asymmetries in the News Impact Curve. *National Centre of Competence in Research Financial Valuation and Risk Management, Working Paper Series*.

Jondeau, E. (2010). Asymmetry in Tail Dependence of Equity Portfolios. *National Centre of Competence in Research, Financial Valuation and Risk Management, Working Paper No. 658*, (No. 658).

Mencía, J. and Sentana, E. (2009). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *Journal of Econometrics*, 153:105–121.

Pearson, K. (1894). Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of Conditional Value at Risk. *Journal of Risk*, 2:21–41.

Part II

Research Papers

Chapter 1

Stable Mixture GARCH Models

Stable Mixture GARCH Models*,†

Simon A. Broda[a] Markus Haas[b]
Jochen Krause[c] Marc S. Paolella[c,d] Sven C. Steude[c]

[a]*Department of Quantitative Economics, University of Amsterdam, The Netherlands*

[b]*Institute for Quantitative Business and Economics Research, University of Kiel, Germany*

[c]*Department of Banking and Finance, University of Zurich, Switzerland*

[d]*Swiss Finance Institute*

Abstract

A new model class for univariate asset returns is proposed which involves the use of mixtures of stable Paretian distributions, and readily lends itself to use in a multivariate context for portfolio selection. The model nests numerous ones currently in use, and is shown to outperform all its special cases. In particular, an extensive out-of-sample risk forecasting exercise for seven major FX and equity indices confirms the superiority of the general model compared to its special cases and other competitors. An improved method (in terms of speed and accuracy) is developed for the computation of the stable Paretian density. Estimation issues related to problems associated with mixture models are discussed, and a new, general, method is proposed to successfully circumvent these. The model is straightforwardly extended to the multivariate setting by using an independent component analysis framework. The tractability of the relevant characteristic function then facilitates portfolio optimization using expected shortfall as the downside risk measure.

Keywords — Density Forecasting; Expected Shortfall; Fat Tails; ICA; GARCH; Mixtures; Portfolio Selection; Stable Paretian Distribution; Value-at-Risk.

JEL Classification: C13, C16, C22, C32, G17.

*Part of the research of Paolella has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation. The research of Haas was supported by the German Research Foundation (DFG).
The authors are grateful to Kerstin Kehrle, Mico Loretan, Benoît Mandelbrot, J. Huston McCulloch, John P. Nolan, Eric Renault, Gennady Samorodnitsky, Murad S. Taqqu, David Veredas, Casper G. de Vries, and two anonymous referees, for providing an array of insightful and detailed comments and suggestions which led to a greatly improved paper.

†A condensed version was published in the Journal of Econometrics, 2013, Vol. 172:2, pp. 292–306.

1.1 Introduction

Starting with the pioneering works of Mandelbrot (1963) and Fama (1965), a variety of studies have investigated the use of the stable Paretian distribution for modeling the unconditional distribution of asset returns. Given the very fat-tailed nature of weekly, daily, and higher-frequency financial returns data, it is not surprising that the stable distribution has been very successful in this regard. While there exist other fat-tailed, asymmetric distributions which also fit asset returns well, (see, e.g., Knight and Satchell, 2001, Paolella, 2007, and the references therein), none of these are closed under summation, a feature which is of great practical use in portfolio allocation; see e.g., Doganoglu et al. (2007) and Giacometti et al. (2007).

It is well-known that the stable Paretian is the only valid distribution that arises as a limiting distribution of sums of independently, identically distributed (iid) random variables. Given that error terms in econometric models are usually interpreted as random variables that represent the sum of all the effects not being captured by the model, the use of the stable Paretian assumption should be highly desirable; for further discussion see McCulloch (1996), Rachev and Mittnik (2000), Rachev (2003) and Nolan (2012).

Despite the advantages of the stable distribution for modeling real phenomena, a popular, albeit often misguided, critique of the use of the stable Paretian distribution is the lack of existence of the second moment. (Recall that unless the tail index $\alpha \in (0, 2]$ equals two, the tails are so heavy that absolute moments of order α and higher do not exist.) A number of studies have attempted to measure the tail index of the distribution of financial returns, as information about the tail index can be used to derive the probability of large price movements and, especially, market crashes (Jansen and de Vries, 1991). Additionally, knowing the maximum existing moment of the return process is of interest, as the lack of second moments will have consequences for risk and portfolio analysis. Nevertheless, it has been demonstrated by several authors that this endeavor is extremely difficult, so that no conclusive evidence of whether second moments of daily financial returns exist or not has been presented. For example, Kratz and Resnick (1996) discuss the inevitable and potentially "outrageous" bias inherent in tail thickness estimators such as the (in)famous Hill estimator (Hill, 1975).

The extreme bias problem for the Hill estimator with stable Paretian distributions has been studied in depth by McCulloch (1997), Mittnik et al. (1998) and Weron (2001). These and related studies have caused a shift from tail estimation to the adoption of a fat-tailed parametric assumption, such as the stable distribution. Indeed, this is embodied in the bold statement of Adler (1997), who states that "*Many of the problems faced by the Hill and related estimators of the tail decay parameter α can be overcome if one is prepared to adopt a more parametric model and assume, for example, stable innovations*". He goes on to say that "*Overall, it seems that the time may have come to relegate Hill-like estimators to the Annals of Not-Terribly-Useful Ideas.*" With the above comments about tail estimation in mind (and Adler's positive stance towards use of the stable distribution), we know of no definitive way to test the existence or nonexistence of second moments. Based on the positive empirical findings below, we conclude that, for applications to density and risk forecasting, the question of existence of second moments is only of secondary interest. This stance is further supported by Malevergne et al. (2005) who state that "*for most practical applications, the relevant question is not to determine what is the true asymptotic tail, but what is the best effective description of the tails in the domain of useful applications*".

Another critique of the use of the stable Paretian distribution is the complexity of computing its density, as required for the likelihood—which for conditional (non-iid) models, is required for parameter estimation. With modern computing power, and the availability of several algorithms, this is no longer a hindrance. Nevertheless, there still appears to be no existing method which is both fast and delivers the high accuracy required for likelihood optimization. To this end, we propose yet another method for its computation. It capitalizes on the vectorized nature of modern computing languages and yields a method which is very fast, but also superior to existing algorithms in terms of accuracy. Its details, and references to other methods (and their flaws) are detailed below in Section 1.2.3.

The real problem with the use of the stable-Paretian, or *any* skewed, fat-tailed distribution for modeling the unconditional distribution of asset returns, is that they cannot capture the time-varying volatility so strongly evident in daily and higher-frequency returns data.[3,4] Section 1.2 discusses the use of the

[3]The term "volatility" is usually defined as the standard deviation of the log returns. In the models we shall be considering, the standard deviation is infinite, so this is strictly speaking a misnomer. In a slight abuse of terminology we shall continue to use the term and take it to refer to the (possibly time-varying) scale of the conditional return distribution.

[4]There is actually another problem with the stable Paretian (or mixtures thereof) which we do not address. The summability (or stability) property of the stable distribution and the definition of log returns implies that the tail index of the return distribution should remain the same at any frequency, i.e., intraday, daily, weekly, monthly, etc.. However, it is well-known that this is usually not the case, with, say, daily returns exhibiting a tail index considerably lower than two, but monthly data exhibiting nearly normal behavior. This occurs because, for such series, the returns are not iid stable Paretian, but rather have a distribution such that, via a central limit theorem, their sums approach normality. The iid aspect of this problem is addressed in Paolella (2001) by accounting for the non-constant scale term by application of a stable-GARCH filter and construction of a formal testing procedure, but even then, the null hypothesis of stability can be rejected for many (but not all) return series. This result is not in conflict with our stable mixture GARCH model because our goal is to (i) devise a model endowed with some plausible statistical and economic motivation, (ii) which yields relatively superior density and risk forecasts for daily (and possibly higher frequency) data, and (iii) can be used in a multivariate context via an ICA decomposition, but without concern for the stability (or lack thereof) aspect of returns. A possible way of incorporating all such features would be to use the tempered stable distribution, which also has a tractable characteristic function, mimics the shape of the stable distribution, but is such that, when iid copies are summed, the tail index increases; see Kim et al. (2008) and Kim et al. (2010).

stable distribution in conjunction with GARCH models to overcome this limitation.

Another popular and successful approach to the unconditional modeling of asset returns and VaR prediction involves the use of finite mixtures of normal distributions. Owing to its great flexibility, a normal mixture, even with just two components, is well-suited for capturing the usual stylized facts typical in a financial context. This model has been motivated and investigated by numerous authors, including Kon (1984), who suggests that returns may be influenced by a series of different information flows including a non-information distribution, a firm-specific information distribution, and a market-wide information distribution—hence, a mixture of three normal distributions. A different economic motivation for the presence of a mixture of distributions is provided by Vigfusson (1997), who builds on theoretical work which explains the stylized facts of financial time series by the interaction of heterogeneous groups of agents, with the groups processing market information differently; see, e.g., Samanidou et al. (2007) for an overview of such models. This is in line with recent research with experimental data by Kirchler and Huber (2007), who show that heterogeneous fundamental information can be a major source for the emergence of fat tails and volatility clustering. The fact that, for each component, a central limit theorem argument can be used to justify the use of the normal distribution is appealing, and lends some theoretical justification for the model and its economic interpretations. The same holds for the use of stable distributions for the mixture components, via the generalized central limit theorem. Indeed, Salas-Gonzalez et al. (2009) propose mixtures of stable distributions with a view towards applications in engineering such as image and radar signal processing.

The aforementioned problem regarding stable distributions also applies to the use of mixtures: in an unconditional setting, the mixture cannot capture the strong time-varying volatility of the returns. This is addressed by several authors who combine mixture models with GARCH structures. In this paper, we propose a model which generalizes the normality assumption in the normal mixture GARCH model to allow for stable distributions, investigate some of its theoretical and empirical properties, and extend its use (in a limited way adequate for portfolio optimization) to a multivariate framework. Moreover, as the estimation of all mixture models, and particularly mixture GARCH models, is numerically challenging due to the degeneracy problem, we introduce new estimators which elegantly resolve this. The degeneracy problem is illustrated graphically using real data, and the excellent performance of the new estimators is demonstrated via simulation.

The remainder of this paper is as follows. Section 1.2 introduces the new mixture GARCH model and the method we suggest for computing the stable density. The new estimators are presented in Section 1.3. Section 1.4 presents an empirical exercise. Section 1.5 details how the model lends itself to portfolio allocation. Section 1.6 concludes.

1.2 Stable Mixture GARCH

Among conditional volatility models, the normal-GARCH has proven itself to be highly effective, though despite its success in capturing a high percentage of the volatility movement, countless applications have confirmed that the residuals, or filtered innovations when applied to weekly, daily, or higher frequency asset return data, still deviate considerably from normality. This has given rise to a large number of alternative models which replace the normal distribution in the GARCH model by a fat-tailed, asymmetric one. Given its theoretical properties, the stable Paretian distribution suggests itself, as first proposed by McCulloch (1985). See Mittnik et al. (2002) for further references and technical details, and Mittnik and Paolella (2003) for a demonstration of its effectiveness in value at risk (hereafter VaR) forecasting.

1.2.1 Mixture GARCH

A model which addresses the fat-tailed, asymmetric innovation issue mentioned above, but also gives rise to rich volatility dynamics not possible in the traditional battery of GARCH models, involves the use of mixtures. Building on the success of the mixed normal distribution for capturing the unconditional skewness and excess kurtosis of asset returns, and on some special cases already in the literature, Haas et al. (2004b) and Alexander and Lazar (2006) independently propose a general model structure which endows each mixed normal component with a GARCH structure.

As in Haas et al. (2004b), we say that time series $\{\varepsilon_t\}$ is generated by a k-component mixed normal GARCH(r, s) process, denoted MixNormal-GARCH, if the conditional distribution of ε_t is a k-component mixed normal distribution with zero mean,

$$\varepsilon_t \mid \mathcal{F}_{t-1} \sim \mathrm{MixNormal}\left(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t\right), \tag{1.1}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and $\boldsymbol{\sigma}_t = (\sigma_{1,t}, \ldots, \sigma_{k,t})'$ are column vectors, the mixed normal probability density function (pdf) is given by

$$f_{\varepsilon_t|\mathcal{F}_{t-1}}\left(x; \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t\right) = \sum_{i=1}^{k} \omega_i \phi\left(x; \mu_i, \sigma_{i,t}\right),$$

$\mathcal{F}_t$ represents the information available at date t, ϕ is the normal pdf, $\omega_i \in (0, 1)$ with $\sum_{i=1}^{k} \omega_i = 1$ and, to ensure zero mean, $\mu_k = -\sum_{i=1}^{k-1} \left(\omega_i/\omega_k\right) \mu_i$. The component variances $\sigma_{i,t}^2$ follow the GARCH-like structure

$$\boldsymbol{\sigma}_t^{(2)} = \boldsymbol{\gamma}_0 + \sum_{i=1}^{r} \boldsymbol{\gamma}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(2)}, \tag{1.2}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i,1}, \gamma_{i,2}, \ldots, \gamma_{i,k})'$, $i = 0, \ldots, r$, are $k \times 1$ vectors, $\boldsymbol{\Psi}_j$, $j = 1, \ldots, s$, are $k \times k$ matrices, $\boldsymbol{\sigma}_t^{(\delta)}$ is short for $(\sigma_{1,t}^{\delta}, \sigma_{2,t}^{\delta}, \ldots, \sigma_{k,t}^{\delta})'$, and $\delta \in \mathbb{R}_{>0}$. The parameters of the model need to be such that $\boldsymbol{\sigma}_t^{(\delta)} > 0$, where, in case of non-scalars, $>$ indicates element-wise inequality. As discussed in the above references, we take $r = s = 1$ and restrict the $\boldsymbol{\Psi}_j$ to be diagonal, so in this case we will refer to the diagonal elements of $\boldsymbol{\Psi}_1$ as ψ_i, $i = 1, \ldots, k$. Further details, and a survey of the model and the extensions which followed are given in Haas and Paolella (2011).

Another feature of mixture GARCH models is their ability to contain conditional as well as unconditional component models. As it turns out, the component of the mixture assigned to the most volatile observations can often be adequately modeled by a component with a relatively high, but constant, variance—it does not require a GARCH structure. We identify such models, where only g, $g \leq k$, components follow a GARCH$(1,1)$ process, by appending the accessory (k, g). Thus, the k-component MixNormal-GARCH model given in (1.1), with $r = s = 1$ and $\boldsymbol{\Psi}_1 = \text{diag}(\psi_1, \ldots, \psi_k)$, is referred to as MixNormal(k, g).

Similar to replacing the normal assumption in a standard GARCH model, it suggests itself to challenge the MixNormal assumption with alternative distributions in addition to, or instead of, increasing the number of mixture components, to achieve better out-of-sample forecasting performance. As demonstrated in Kuester et al. (2006), one or more of the component densities of the MixNormal model can still exhibit tails which are fatter than the normal. In Kuester et al. (2006) and Rombouts and Bouaddi (2009) the generalized exponential distribution (in short, GED) is applied and shown to lead to improved in-sample fits and quality of VaR forecasts. For comparison with the MixStable-GARCH model, and using the aforementioned (k, g) notation, we will refer to the MixGED-GARCH model as MixGED(k, g) in the following.

In our empirical section, we use the GED distribution given by the location-zero, scale-one pdf

$$f(x; p) = \frac{p}{2\Gamma(p^{-1})} \exp\{-|x|^p\}, \quad p > 0.$$

Its cumulative distribution function (cdf), required for VaR calculations, follows as

$$F(x; p) = \frac{1}{2}\left(1 - \bar{\Gamma}_{(-x)^p}\left(p^{-1}\right)\right), \quad x \leq 0,$$

where $\bar{\Gamma}$ is the incomplete gamma ratio, and, for $x > 0$, $F(x) = 1 - F(-x)$ due to symmetry.

1.2.2 The MixStable Model

Analogous to (1.1) and (1.2), we say that time series $\{\varepsilon_t\}$ follows a k-component mixed stable GARCH(r, s) process, denoted MixStable-GARCH, if the distribution of $\varepsilon_t \mid \mathcal{F}_{t-1}$ is a finite mixture of stable distributions. Its pdf is

$$f_{\varepsilon_t|\mathcal{F}_{t-1}}(x; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t) = \sum_{i=1}^{k} \omega_i f_S(x; \alpha_i, \beta_i, \mu_i, \sigma_{i,t}), \tag{1.3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$ is the set of tail indices, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is the set of asymmetry parameters corresponding to the k stable distributional components, and, as before, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)'$ is the set of weights, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ is the set of component location terms, $\boldsymbol{\sigma}_t = (\sigma_{1,t}, \ldots, \sigma_{k,t})'$ is the set of strictly positive scale parameters, and $f_S(x; \alpha, \beta, \mu, \sigma)$ is the location-μ, scale-σ, stable Paretian pdf with tail index α and skewness parameter β, as in Samorodnitsky and Taqqu (1994). For the mixture GARCH, let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)'$ be the set of power GARCH coefficients. Further, we assume that $1 < \alpha_i \leq 2$, so that the mean exists, and, with $\alpha_{\min} = \min_i \alpha_i$, restrict $0 < \delta_i < \alpha_{\min}$, $i = 1, \ldots, k$. This is a natural extension of the power restriction in the stable-GARCH model as devised in Mittnik et al. (2002). If $\alpha_{\min} = 2$ (so that $\alpha_i = 2$ for all i), the δ_i need only be positive, $i = 1, \ldots, k$. As with MixNormal(k, g), to ensure zero mean, $\mu_k = -\sum_{i=1}^{k-1} (\omega_i/\omega_k)\, \mu_i$ is imposed. The component scale terms, analogous to the variance term in the MixNormal model, evolve according to

$$\boldsymbol{\sigma}_t^{(\boldsymbol{\delta})} = \boldsymbol{\gamma}_0 + \sum_{i=1}^{r} \boldsymbol{\gamma}_i \left|\varepsilon_{t-i}\right|^{(\boldsymbol{\delta})} + \sum_{j=1}^{s} \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(\boldsymbol{\delta})}, \tag{1.4}$$

where $\boldsymbol{\sigma}_t^{(\boldsymbol{\delta})}$ is short for $(\sigma_{1,t}^{\delta_1}, \sigma_{2,t}^{\delta_2}, \ldots, \sigma_{k,t}^{\delta_k})'$. Motivated by our use of relatively (for GARCH applications) small sample sizes (see the comment in Footnote 13 and the discussion in Section 1.4.3 below), we impose $\alpha_i = \alpha_j$ and $\beta_i = \beta_j$ for all stable mixture components as well as $\delta_i = \delta_j$ for all mixture GARCH models in the following. As such, we drop the component index i and just write α, β and δ. Similar to MixNormal(k, g), MixStable(k, g) denotes the k-component mixed stable GARCH$(1, 1)$ process with diagonal $\boldsymbol{\Psi}_1$ matrix, and only g of the k components having a GARCH structure. For $\beta = 0$, the model nests the following:

1) The unconditional stable Paretian model, as proposed by Mandelbrot (1963) and Fama (1965), by taking $k = 1$ and no GARCH structure.

2) The unconditional mixed normal model from Fama (1965), Kon (1984) and others, by taking $\alpha_i =$ in each of the k components, and no GARCH structure.

3) The normal-GARCH model from Bollerslev (1986), by taking $k = 1$, $\alpha = 2$ and $\delta = 2$.

4) The symmetric stable-GARCH model of Mittnik et al. (2002), by taking $k = 1$.

5) The MixNormal(k, g) model (1.1) and (1.2) of Haas et al. (2004b) and Alexander and Lazar (2006), by taking $\delta = 2$ and $\alpha = 2$.

6) The "linear" two-component MixNormal-GARCH models of Vlaar and Palm (1993) and Bai et al. (2003).

The MixStable-GARCH model is very general; in particular, there are two sources of asymmetry in the model: μ_i and β. In the empirical section below, we will limit ourselves to two special cases, which we dub A^{1}MixStable(k, g) and A^{2}MixStable(k, g), respectively. In the former, we restrict β to zero, whereas in the latter, $\mu_i = 0$, so that only one source of asymmetry is considered at a time.

As the stable distribution does not possess a finite second moment if $\alpha < 2$, the MixStable-GARCH process will not be covariance stationary. It may still be strictly stationary, however. This follows from Liu (2007) who generalizes the Markov-switching GARCH(1,1) process of Haas et al. (2004a) to allow for more general power parameters and innovation distributions, and studies its dynamic properties. This process nests the model studied in this paper, and we state the stationarity condition for the parsimonious case where all the stable components are characterized by the same shape parameters α and β. It then follows from Corollary 2.1 of Liu (2007) that a sufficient condition for the process to be strictly stationary with a finite δth moment is that the eigenvalues of the matrix

$$\boldsymbol{\gamma}_1 \boldsymbol{\omega}' \kappa_{\delta,\alpha,\beta} + \boldsymbol{\Psi}_1 \tag{1.5}$$

are inside the unit circle, where, as in Mittnik et al. (2002),

$$\begin{aligned}
\kappa_{\delta,\alpha,\beta} &= \eta_\delta^{-1} \Gamma\left(1 - \frac{\delta}{\alpha}\right)\left(1 + \tau_{\alpha,\beta}^2\right)^{\delta/(2\alpha)} \cos\left(\frac{\delta}{\alpha} \arctan\left(\tau_{\alpha,\beta}\right)\right), \qquad (1.6) \\
\tau_{\alpha,\beta} &= \beta \tan\left(\alpha\pi/2\right), \\
\eta_\delta &= \begin{cases} \Gamma\left(1-\delta\right)\cos\left(\frac{\pi\delta}{2}\right), & \text{if } \delta \neq 1, \\ \pi/2, & \text{if } \delta = 1. \end{cases}
\end{aligned}$$

Term $\kappa_{\delta,\alpha,\beta}$ is the power-δ absolute moment of a stable random variable with tail index α and asymmetry parameter β; see Paolella (2007, Sec. 8.3) for a detailed derivation. For $\delta = 1$ and $\beta = 0$, (1.6) reduces to $\kappa_{1,\alpha,0} = 2\,\Gamma\left(1 - \alpha^{-1}\right)/\pi$.

Similar to the mixed normal GARCH model, condition (1.5) allows some (but not all) components to be driven by non-stationary GARCH dynamics, whereas overall the process will be stationary as long as the mixing weights of these components are sufficiently small. In particular, when $\max_j\{\psi_j\} < 1$ is satisfied, it follows from arguments similar to Haas et al. (2004a) that the eigenvalue condition is equivalent to

$$\sum_{j=1}^{k} \omega_j \frac{\kappa_{\delta,\alpha,\beta}\gamma_{1,j}}{1 - \psi_j} < 1. \tag{1.7}$$

The ARCH(∞) representation of $\sigma_{j,t}^\delta$, given by

$$\sigma_{j,t}^\delta = \frac{\gamma_{0,j}}{1 - \psi_j} + \gamma_{1,j} \sum_{i=1}^{\infty} \psi_j^{i-1} |\varepsilon_{t-i}|^\delta,$$

shows that the total impact of a shock on future volatility in regime j is

$$\gamma_{1,j} \sum_{i=0}^{\infty} \psi_j^i = \frac{\gamma_{1,j}}{1-\psi_j}.$$

Thus, condition (1.7) restricts the *average* total impact of a shock on the component-specific future volatilities.

1.2.3 Computational Remarks

From (1.3) and (1.4), it is readily apparent that the likelihood of the MixStable model is straightforward to calculate, provided a computable expression for the density of the stable Paretian distribution is available. Several authors have developed methods for this, including Doganoglu and Mittnik (1998), McCulloch (1998), Nolan (1998) and Mittnik et al. (1999). For evaluating (1.3), we compute the stable densities of the mixture components based on a variant of the real-valued integral expression of Zolotarev (1986) as given in Nolan (1997).[5]

1.2.3.1 Evaluation of the Stable Density

Considering existing first moments ($\alpha > 1$), the original expression of the stable pdf in Nolan (1997) reduces to

$$f_S(x, \alpha, \beta) = \frac{1}{\sigma} \begin{cases} \frac{\alpha}{\pi|\alpha-1|(z-\zeta)} \int_{-\tau}^{\pi/2} V(y;\alpha,\beta,z) \exp\{-V(y;\alpha,\beta,z)\}\, \mathrm{d}y, & \text{if } z > \zeta, \\ \Gamma(1+\frac{1}{\alpha})/\pi \cos(\tau)(1+\zeta^2)^{-1/(2\alpha)}, & \text{if } z = \zeta, \\ f_S(-z, \alpha, -\beta), & \text{if } z < \zeta, \end{cases} \tag{1.8}$$

where z is the standardized and transformed observation, $z = (x-\mu)/\sigma - \beta\tan(\alpha\pi/2)$, and

$$\begin{aligned} V(y;\alpha,\beta,z) &= \cos(\alpha\tau)^{1/(\alpha-1)} \left(\frac{(z-\zeta)\cos(y)}{\sin(\alpha y + \alpha\tau)}\right)^{\alpha/(\alpha-1)} \left(\frac{\cos(\alpha\tau + y(\alpha-1))}{\cos(y)}\right), \\ \zeta &= -\beta \tan\left(\frac{\pi\alpha}{2}\right), \\ \tau &= \frac{1}{\alpha} \arctan\left(\beta \tan\left(\frac{\pi\alpha}{2}\right)\right). \end{aligned}$$

For the fast computation of several evaluation points at once a vectorized implementation of (1.8) suggests itself, so we replace all mathematical operators by their element-wise counterparts. In doing so, we adapt a vectorized variant of the adaptive Simpson quadrature, (e.g., see Lyness, 1969) given by

$$\oint\!\!\!\!\square_{\boldsymbol{a}}^{\boldsymbol{b}} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \begin{cases} (\boldsymbol{q}_1 + \boldsymbol{q}_2)/2, & \text{if } \|\boldsymbol{q}_2 - \boldsymbol{q}_1\|_\infty \le 10^{-6}, \\ \oint\!\!\!\!\square_{\boldsymbol{a}}^{\boldsymbol{c}} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \oint\!\!\!\!\square_{\boldsymbol{c}}^{\boldsymbol{b}} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, & \text{otherwise}, \end{cases} \tag{1.9}$$

[5] We re-scale the distribution by $1/\sqrt{2}$, so that the standard normal distribution arises as a special case for $\alpha = 2$, or similarly for $p = 2$ in the GED case.

where $\oint$ denotes the (recursive) Simpson integral,

$$\boldsymbol{q}_1 = \frac{\boldsymbol{b}-\boldsymbol{a}}{6}\left(f(\boldsymbol{a})+4f(\boldsymbol{c})+f(\boldsymbol{b})\right) \text{ and } \boldsymbol{q}_2 = \frac{\boldsymbol{b}-\boldsymbol{a}}{12}\left(f(\boldsymbol{a})+4f(\boldsymbol{d})+2f(\boldsymbol{c})+4f(\boldsymbol{e})+f(\boldsymbol{b})\right),$$

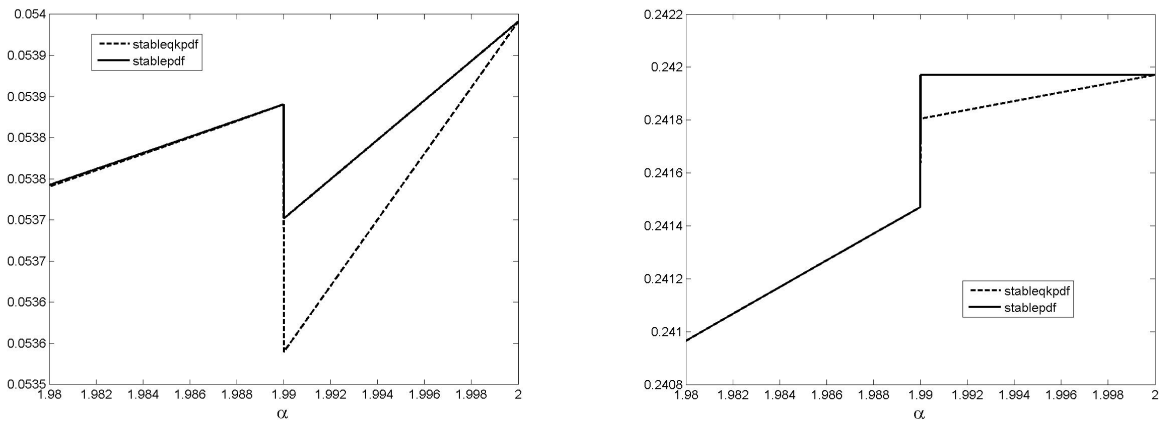
as well as $\boldsymbol{c} = (\boldsymbol{a}+\boldsymbol{b})/2$, $\boldsymbol{d} = (\boldsymbol{a}+\boldsymbol{c})/2$ and $\boldsymbol{e} = (\boldsymbol{c}+\boldsymbol{b})/2$. Besides the reduced computational overhead by virtue of vectorization, the computation is further accelerated by exploiting redundancies. For fixed α and β as well as location-zero scale-one z, observe that several parts of (1.8) remain constant and only need to be evaluated once for all z. By removing these redundancies, we find that computation times decrease significantly, in particular those of the integrand, since the recursive Simpson quadrature requires a large number of function evaluations (a fact that is shared by most numerical integration methods). The resulting routine for the stable Paretian density is about 40 times faster than the naive implementation but equally robust and accurate. Moreover, the routine is also 8 times faster than the direct evaluation of the stable density found in John Nolan's STABLE 4.0 toolbox. In addition, that routine has an error, see Figure 1.

1.2.3.2 Further Methods of Density Approximation

As mentioned above, several ways have been proposed for computing the density of a stable Paretian variate. In addition to the direct approach in the previous section based on the real integral representation, we consider two further methods: First, the fast Fourier transform (FFT) based inversion of the characteristic function; and second, the spline-based, very fast approximation developed for the STABLE toolbox of John Nolan.

The FFT approach first inverts the characteristic function based on a finite grid and then evaluates the pdf at the desired points by a subsequent interpolation. This approach is generally applicable and quite effective whenever the characteristic function is easily evaluated; see Mittnik et al. (1999); and Paolella (2007, Chapters 1 and 8). The accuracy of the approach (controlled by the grid size), however, is limited as the maximal grid size is restricted by memory constraints. For extreme events in the outer tail area in particular, the achievable accuracy in practice is not sufficient for likelihood optimizations as the routine may return the same likelihood at different evaluation points. Besides, there are estimation issues with dynamic grid sizes, which are often used for speed improvements. What turns out to be a convenient mechanism for reducing computation times becomes problematic in likelihood optimization as likelihood values become a function of the (varying) grid size. As a consequence, the estimation may not converge or, even worse, the estimation problem might not be identified (numerically) anymore.

In contrast, turning to Nolan's STABLE 4.0 toolbox (being closed source unfortunately), the spline approximation therein is more robust and, as to be expected for approximations, much faster. The increase in execution time is roughly a factor of 500 compared to the vectorized version of (1.8), though when computing the likelihood of the MixStable-GARCH model, this factor is only about 3 to 5, because of the time required for the matrix-based mixture GARCH filter. However, by construction, spline approximations will be correct to only a few significant digits, and can sometimes be problematic. Unfortunately, we also find numerical inconsistencies for $\alpha > 1.99$ as an error source in our estimations, see Figure 1.1. To circumvent these problems we rely on alternative methods (e.g., interpolation) in this region.



$x = 2, \beta = 0.2, \mu = 0, \sigma = 1$ $\qquad$ $x = 0, \beta = 0, \mu = 1, \sigma = 1$

Figure 1.1: Plots of the stable Paretian density as a function of the α-stable parameter in the neighborhood of $\alpha = 1.99$ for both the spline approximation, `stableqkpdf`, and the direct evaluation, `stablepdf`, as found in Nolan's STABLE 4.0 toolbox. Both panels illustrate a numerical discontinuity. This discontinuity indeed causes problems during optimization. For illustration, the remaining parameters are fixed to the values shown below each panel, but the problem arises for different values of x, β, μ and σ as well. Moreover, the problem is also found in the corresponding cdf routines.

Our tests, based on real and simulated data, using (1.9) and Nolan's spline approximation suggest that the best compromise between speed and accuracy is a combination of both. As the fastest method, the spline approximation (with correction for the discontinuity mentioned above) lends itself well for an initial exploration of the search space, while the direct evaluation of (1.8) yields more reliable and accurate results. As such, we use Nolan's spline approximation to obtain initial estimates for the MixStable-GARCH model, which are then used as starting values in the subsequent re-estimation of the model based on the vectorized routine for (1.8). Results only based on the spline approximation were clearly inferior to those which are "polished" with a more accurate evaluation of the density.

1.2.3.3 Quantiles

In the empirical analysis below, we will additionally need the quantiles of the distribution from (1.3) in order to compute the VaR. If X follows a stable mixture random variable with pdf as in (1.3), then its cdf is just $\Pr(X \leq x) = \sum_{i=1}^{k} \omega_i F_S(x; \alpha_i, \beta_i, \mu_i, \sigma_i)$, where $F_S(x; \alpha, \beta, \mu, \sigma)$ is the cdf of the stable. The latter can be evaluated using the Gil-Pelaez (1951) inversion formula or the cdf integral expression of Zolotarev (1986) which avoids the need for complex numbers. For the case with $\alpha > 1$, Zolotarev's cdf expression is given by

$$F_S(z, \alpha, \beta) = \begin{cases} 1 - \frac{1}{\pi} \int_{-\tau}^{\pi/2} \exp\{-V(y; \alpha, \beta, z)\} \, \mathrm{d}y, & \text{if } z > \zeta, \\ \left(\frac{\pi}{2} - \tau\right)/\pi, & \text{if } z = \zeta, \\ 1 - F_S(-z, \alpha, -\beta), & \text{if } z < \zeta. \end{cases}$$

1.2.3.4 Optimization and Starting Values

Regarding the maximization of mixture likelihood functions, it appears beneficial to diversify among existing optimization methods for improving the chances of finding the global maximum given the inherently bumpy surface of such functions. We consider a combined optimization approach based on different search strategies from the rich set of unconstrained optimization techniques, including quasi-Newton and simplex methods. Using unconstrained optimization techniques, all constraints must be satisfied manually. We compute the mixture weights and the location parameters, respectively, from the $k-1$ estimated coefficients, by

$$\omega_i = \begin{cases} \hat{\omega}_i \left(1 - \sum_{j=1}^{i-1} \omega_j\right), & \text{if } i < k, \\ 1 - \sum_{j=1}^{k-1} \omega_j, & \text{if } i = k, \end{cases} \quad \text{and} \quad \mu_i = \begin{cases} \hat{\mu}_i, & \text{if } i < k, \\ -\left(\sum_{j=1}^{k-1} \omega_j \mu_j\right)/\omega_k, & \text{if } i = k, \end{cases}$$

where $\hat{\mu}_i$ and $\hat{\omega}_i$ are supplied by the optimization. Alternatively, $\boldsymbol{\omega}$ can be computed from the k coefficients by $\omega_i = \hat{\omega}_i / \sum_{j=1}^{k} \hat{\omega}_j$.[6] Box-constraints on single parameters, e.g., $1 < \alpha_i \leq 2$ and $0 \leq \omega_i \leq 1$, are satisfied by simple interval transformations. All estimations are terminated at a maximum of $50,000$ function evaluations, or, whenever the maximal change in the function value or the parameter vector is smaller than 10^{-4}. Numerical problems due to $\sigma_{i,t} \to 0$ or $\sigma_{i,t} \to \infty$ are addressed by enforcing $\underline{x} \leq \sigma_{i,t} \leq \overline{x}$, where $\underline{x}$ and $\overline{x}$, respectively, denote the smallest and largest (finite) machine number.

In addition to the above optimization technique and the joint estimation procedure in Section 1.2.3.2, we further improve upon the estimation quality by repeating the procedure with different starting values, and picking the best outcome in terms of the (augmented) likelihood value. Along with use of starting

[6] This, however, is numerically less efficient as it unnecessarily increases the dimension of the problem.

values based on previously estimated financial returns series, we also use values which are uniformly drawn from the allowed parameter space. For each estimation, at least three random starting vectors are considered. In consecutive estimations (such as used in our out-of-sample forecast exercises) we also include the previous best estimate. In all situations, the proposed estimation procedure converged irrespective of the starting value. In general, this would be remarkable, given the problematic nature of mixture likelihood functions, but is due to the use of the ALE estimation method discussed below in Section 1.3. Moreover, often (but not always), the various starting values led to the same optimum, further increasing our confidence that it is the global maximum.

1.3 Estimation of Mixture GARCH Models

It is well-known that the likelihood function of a mixture is potholed with singularities (infinite likelihood values), where single mixture components approach Dirac's delta distribution (see Kiefer and Wolfowitz, 1956; and Day, 1969). In the estimation of MixNormal-GARCH models, such *degenerated* states can be avoided by using Bayesian estimation procedures as devised in Ausín and Galeano (2007) and Bauwens et al. (2007). A potential drawback of these approaches, however, is their computational complexity and relatively high estimation time, along with the added complexity of augmenting those procedures from the normal to the stable (or other) distribution. To address these shortcomings, we devise an augmented likelihood function which can still be maximized with conventional optimization techniques but, unlike the usual likelihood, completely and elegantly avoids degenerated mixture estimates. As the problem of avoiding degenerated mixture estimates is a general problem known by many names, we will refer to it as the *mixture degeneracy problem*.

1.3.1 Augmented Likelihood Estimation

Augmented likelihood estimation (ALE) is based on the observation that degenerated components yield zero likelihood values for all observations which differ from their location parameter. As such, singularities can be prevented by avoiding zero likelihood values in all mixture components. This idea is implemented in the augmented likelihood framework by adding the geometric average likelihood value of each mixture component to the overall likelihood function. The additional k likelihood terms yield an infinite penalty whenever a mixture component degenerates. Thus, the ALE solves the mixture degeneracy problem without parameter constraints or penalty terms, but rather by shrinking the overall mixture likelihood function towards the component-wise likelihood functions.

Consider first estimation of an unconditional (no GARCH) mixture distribution. Let $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k)'$

denote the vector of model parameters, where $\boldsymbol{\omega}$ is the vector of mixture weights, and $\boldsymbol{\theta}_i$ represents the parameter vector of the density function, f_i, of the ith component. Then, the ALE takes the form of $\hat{\boldsymbol{\theta}}_{\text{ALE}} = \arg\max_{\boldsymbol{\theta}} \tilde{\ell}\left(\boldsymbol{\theta}; \boldsymbol{\varepsilon}\right)$, where the augmented log-likelihood function $\tilde{\ell}$ consists of the usual sum of log-likelihood values $\ell^{\star}$, as well as the logarithmic geometric means of the component-wise likelihood series $\bar{\ell}_i$, i.e.,

$$\begin{aligned}\tilde{\ell}\left(\boldsymbol{\theta}; \boldsymbol{\varepsilon}\right) &= \ell^{\star}\left(\boldsymbol{\theta}; \boldsymbol{\varepsilon}\right) + \sum_{i=1}^{k} \bar{\ell}_i\left(\boldsymbol{\theta}_i; \boldsymbol{\varepsilon}\right) \\ &= \sum_{t=1}^{T} \log \sum_{i=1}^{k} \omega_i f_i\left(\varepsilon_t; \boldsymbol{\theta}_i\right) + \sum_{i=1}^{k} \frac{1}{T} \sum_{t=1}^{T} \log f_i\left(\varepsilon_t; \boldsymbol{\theta}_i\right). \end{aligned} \tag{1.10}$$

Assuming (as we subsequently do) that the true mixture process is free of degenerated components and its standard likelihood function has a consistent root (compare Kiefer, 1978), the ALE is consistent, because, by dividing by T, the contribution to $\tilde{\ell}$ from the additional k terms becomes negligible in the limit as $T \to \infty$ for all mixtures without degenerated components.[7]

1.3.2 Augmented Mixture GARCH Estimation

The augmented likelihood methodology above is tailored to unconditional mixtures and, thus, cannot prevent singularities in (time-) conditional mixture models. In k-component mixture GARCH models, for instance, singularities occur at all time points t, $t = 1, \ldots, T$, and all components i, $i = 1, \ldots, k$, where $\sigma_{i,t} = 0$ and $\mu_i = \varepsilon_t$. We refer to this problem as the *local degeneracy problem*. Assuming that the true mixture GARCH process does not possess singular or near-singular components at any point in time, we devise two solutions for the local degeneracy problem based on the augmented likelihood methodology.

The first solution is built upon the ALE by removing singularities from the (mixture) likelihood function using a lower bound on the GARCH constant(s),

$$\gamma_{i,0} > c > 0, \quad \text{for all } \ i = 1, \ldots, k. \tag{1.11}$$

This parallels the idea in Hathaway (1985) in the simple (unconditional) normal mixtures, and is extended in Tanaka (2009). Without local degeneracy, the ALE remains feasible and the mixture GARCH degeneracy problem can be solved as in the unconditional case. The *restricted* ALE (RALE) for the MixStable-GARCH model is given by $\hat{\boldsymbol{\theta}}_{\text{RALE}} = \arg\max_{\boldsymbol{\theta}} \tilde{\ell}\left(\boldsymbol{\theta}; \boldsymbol{\varepsilon}\right)$ where (1.11) is satisfied and (1.10)

[7] An example of a non-consistent ALE (or MLE) can easily be constructed by imposing inappropriate parameter constraints. Indeed, this is the reason why the RALE (below) is not always a consistent estimator.

becomes

$$\tilde{\ell}_S(\boldsymbol{\theta};\boldsymbol{\varepsilon}) = \sum_{t=1}^{T} \log \sum_{i=1}^{k} \omega_i f_S(\varepsilon_t; \alpha_i, \beta_i, \mu_i, \sigma_{i,t}) + \sum_{i=1}^{k} \frac{1}{T} \sum_{t=1}^{T} \log f_S(\varepsilon_t; \alpha_i, \beta_i, \mu_i, \sigma_{i,t}).$$

A shortcoming of this approach is the necessity of having to choose the tuning parameter c, and doing so in such a way that, as a function of sample size T, the sequence c_T yields a consistent estimator. Extensive studies with the financial data used in this report, and with simulated data mimicking such finance data, yield that $c = 0.01$ works well in the sense that, while (1.11) is occasionally binding, the parameter estimates barely differ from those obtained by the EALE method discussed next.

Our second solution to the local degeneracy problem is more general and can be applied to any conditional mixture model. Let $\boldsymbol{\ell} \in \mathbb{R}^T$ denote a vector of (finite) log-likelihood values. Based on the observation that the sample geometric mean of the likelihood values

$$\widehat{\mathbb{E}}_{\text{geo}}(\boldsymbol{\ell}) = \exp\left\{\frac{1}{T}\sum_{t=1}^{T} \ell_t\right\} = \exp\left\{\bar{\ell}\right\}$$

grows at a slower rate than the sample variance-like quantity

$$\widehat{\mathbb{V}}_{\text{geo}}(\boldsymbol{\ell}) = \frac{1}{T}\sum_{t=1}^{T} \left(\exp\{\ell_t\} - \exp\left\{\bar{\ell}\right\}\right)^2$$

if some likelihood values move towards infinity, i.e.,

$$\frac{\widehat{\mathbb{E}}_{\text{geo}}(\boldsymbol{\ell})}{\widehat{\mathbb{V}}_{\text{geo}}(\boldsymbol{\ell})} \to 0\ , \quad \text{if } \widehat{\mathbb{E}}_{\text{geo}}(\boldsymbol{\ell}) \to \infty,$$

we propose the component-wise penalty term, $\check{\ell}_i$,

$$\check{\ell}_i(\boldsymbol{\theta}_i;\boldsymbol{\varepsilon}) = \log\left(1 + \frac{1}{T}\sum_{t=1}^{T}\left(\exp\{\ell_{i,t}\} - \exp\left\{\bar{\ell}_i\right\}\right)^2\right), \tag{1.12}$$

where $\ell_{i,t} = \log f_i(\varepsilon_t;\boldsymbol{\theta}_i)$. The component-wise incorporation of (1.12) into (1.10) consequently removes all local singularities. Moreover, as (1.12) is purely likelihood-based, the resulting estimator is also free of direct parameter constraints. The *extended* ALE (EALE) is given by $\hat{\boldsymbol{\theta}}_{\text{EALE}} = \arg\max_{\boldsymbol{\theta}} \tilde{\tilde{\ell}}(\boldsymbol{\theta};\boldsymbol{\varepsilon})$ where

$$\tilde{\tilde{\ell}}(\boldsymbol{\theta};\boldsymbol{\varepsilon}) \;=\; \ell^{\star}(\boldsymbol{\theta};\boldsymbol{\varepsilon}) \;+\; \sum_{i=1}^{k}\bar{\ell}_i(\boldsymbol{\theta}_i;\boldsymbol{\varepsilon}) \;-\; \sum_{i=1}^{k}\check{\ell}(\boldsymbol{\theta}_i;\boldsymbol{\varepsilon}).$$

Similar to the ALE, the additional terms vanish as $T \to \infty$, so that the EALE is consistent. The price to pay is that the penalty in (1.12) introduces a source of bias as the density function is penalized for (too) large likelihood values, thus implicitly transferring mass to the tail area. For the case of the MixStable-GARCH model, the expression specializes to

$$\tilde{\tilde{\ell}}_S(\boldsymbol{\theta};\boldsymbol{\varepsilon}) = \sum_{t=1}^{T}\log\left(\sum_{i=1}^{k}\omega_i L_{i,t}^S\right) + \sum_{i=1}^{k}\left\{\log g_i - \log\left(1 + \frac{1}{T}\sum_{t=1}^{T}\left(L_{i,t}^S - g_i\right)^2\right)\right\},$$

where $L_{i,t}^S = f_S(\varepsilon_t; \alpha_i, \beta_i, \mu_i, \sigma_{i,t})$ and $g_i = \left(\prod_{t=1}^T L_{i,t}^S\right)^{1/T}$.

Both RALE and EALE yielded essentially the same maximized likelihood values in all runs we conducted based on real and simulated data. (It is important to emphasize that the actual, and not the augmented, likelihood is being referred to here.) While there was a slight preference for the EALE over the RALE in terms of actual likelihood value, both resulted in essentially identical out-of-sample forecasts. The local degeneracy problem for MLE, RALE and EALE is further illustrated in Figure 1.2. For all data generating processes we have studied, the results look qualitatively the same: while RALE and EALE never yield degenerated estimates (all likelihood values are finite), the standard MLE frequently results in degenerated estimates (infinite likelihood values). It is due to the imperfection of global optimization methods that, first, the MLE does not always degenerate, and second, that degenerated estimates occur less frequently for larger sample sizes, as can be seen by comparing the left and right panels in Figure 1.2.

1.3.3 Small Sample Properties

Simulation studies were conducted in order to assess the distributional properties of the estimators, partly to compensate for the fact that the asymptotic properties are elusive at present. The parameters of the data generating processes we consider are calibrated to actual finance data. The first analysis is based on the estimation error $\epsilon_{i,j}^P = \hat{\theta}_{i,j} - \theta_j$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, where $\boldsymbol{\theta}$ is the parameter vector of the true process, M is the total number of parameters, and $N = 1000$ the number of simulations. To address the label switching problem, parameter vectors are sorted by mixture weights. Table 1.1 shows the results of two simulation studies for the MixStable(k, g) model.[8] Two measures are reported. The first is the usual root mean squared error of the parameters (P-RMSE), $\sqrt{\sum_i (\epsilon_{i,j}^P)^2 / N}$.[9] The second measure is the inter-quantile range (IQR).[10] As expected, all error measures decrease as the sample size increases.

We now turn to the second analysis. The above parameter error diagnostic needs to be augmented as a basis for the analysis for two reasons: First, mixture processes can often be mimicked quite accurately by processes with fairly different parameter vectors. Second, the sorting of mixture components may lead to false results, e.g., if some of the true mixture weights are close. Hence, we consider two proxy measures that reflect the main characteristics of mixture GARCH processes. The first is based on the quantiles of

[8] More detailed results and information on computational aspects are available from the authors upon request.

[9] We exclude those runs (estimated parameter vectors) such that one or more components had an excessively small weight (in particular, when $\omega_i T < 10$, e.g., the ith component explains less than 0.1% of the observations when using $T = 1000$). We do this to account for the RMSE sensitivity to numerical outliers in the parameter estimates of such components.

[10] Unlike the P-RMSE, the IQR is based on all runs, as it is not affected by outliers.

the unconditional distribution of the process and the second uses the autocorrelation function (acf) of the absolute returns. (In order to compare the sample values to their theoretical counterparts, the latter are obtained via simulation of the true process using one million observations.) The corresponding RMSE measures (Q- and A-RMSE) are given by the square root of the mean squared

$$\epsilon^Q_{i,j} = \text{quantiles}(\hat{\theta}, \boldsymbol{\zeta})_{i,j} - \text{quantiles}(\theta, \boldsymbol{\zeta})_{0,j} \text{ and } \epsilon^A_{i,j} = \text{acf}(\hat{\theta}, \rho)_{i,j} - \text{acf}(\theta, \rho)_{0,j},$$

respectively, where we evaluate the sample quantile function at the probability levels $\boldsymbol{\zeta} = (0.01, 0.1, 0.2, \ldots, 0.9, 0.99)$ and compute the sample acf with lag order $\rho = 100$.

For comparison we also report the performance of the standard MLE (although it is ill-defined for mixtures). As expected, the MLE often results in estimates for which at least one component explains only a few observations. Such mixture components are typically either degenerated or correspond to local optima where fewer mixture components explain the data (in the sense of almost zero-valued mixture weights) than actually being estimated. Clearly, RALE and EALE outperform the MLE here, but also suffer from (non-singular) local optima, though substantially less. Similarly, it is evident that the EALE results in non-stationary estimates more often than the RALE. The reason is that in order to match a given unconditional volatility, an increase in the GARCH constant must be offset by a decrease in the remaining GARCH parameters, thus moving them away from the non-stationarity border. This shortcoming of the EALE, however, vanishes as the sample size increases. In contrast, by looking at the IQR values based on all estimates, Table 1.1 indicates that RALE and EALE perform very similarly; in the two-component case the RALE slightly outperforms the EALE, while it is the other way around in the three-component case. Although the RALE tends to be slightly faster, we opt for the EALE in the following, as it has the appealing advantage of avoiding direct parameter constraints and is therefore an unrestricted solution of the mixture degeneracy problem.[11]

1.4 Univariate Empirical Results

Our empirical analysis covers the major international equity indices DAX 30, S&P 500, DJIA 30, NIKKEI 225 and NASDAQ COMPOSITE (20 years, dating back from July 7th, 2009; resulting in a sample size of 2609) as well as the exchange rates JPY/EUR and USD/EUR (10 years, dating back from July 7th, 2009; resulting in a sample size of 1304). All results are based on percentage log returns, $\varepsilon_t =$

[11]Regarding computation time: On an Intel i7-2600K quad-core processor at 4.2Ghz a single estimation (EALE) of the MixStable$(4, 4)$ model based on 1000 data points (running in Matlab R2010a on a single core of the cpu) takes up to four minutes using Nolan's fast spline approximation and about four times the amount of time using the vectorized version of Zolotarev's integral expression.

$100\,(\log p_t - \log p_{t-1})$, where p_t is the daily closing index price at time t. We study MixNormal(k, g), MixGED(k, g) and MixStable(k, g) models, where $(k, g) \in \{(2, 2), (3, 2), (3, 3), (4, 4)\}$.[12] To prevent overfitting, shape parameters are restricted to be identical.[13]

Cutting to the chase, the best performing models (in terms of both in-sample fit and, more importantly, out-of-sample performance) are all in the MixStable class. In-sample, the A^{2}MixStable($2, 2$) model is favored by the BIC, while out-of-sample, the A^{2}MixStable($2, 2$), as well as the A^{2}MixStable($3, 2$) and A^{2}MixStable($3, 3$), perform well. It is noteworthy that the A^{1}MixStable($4, 4$) model, despite its high parametrization, performs overall best in terms of the uniformity of the predictive cdf values. Nevertheless, the A^{2}MixStable($3, 2$) model might be preferred because of its relatively parsimonious parametrization (which also implies faster estimation). It also performed slightly better than the A^{1}MixStable($4, 4$) for the VaR comparisons at the lower probability levels.

1.4.1 Choice of δ

We investigate the influence of the GARCH power parameter δ, for which $\delta < \alpha$ must be satisfied in order to ensure the existence of the MixStable-GARCH process. In particular, given the flexibility and richness of the model, we have confirmed for numerous data sets and sample sizes that, relative to the other parameters, the likelihood is relatively flat in δ. This can lead to exacerbated estimation problems (in addition to requiring dynamically imposed constraints, as opposed to simple box constraints, during estimation), and implies that just setting δ to a compromise value will not lead to appreciably poorer (and could possibly lead to slightly better) forecasts. We find that the choice of $\delta = 1$ is not only adequate, but also conveniently satisfies the $\delta < \alpha$ constraint.

Consider the profile log-likelihood function, $\ell_{\mathrm{p}}\left(\delta; \boldsymbol{\varepsilon}\right) = \max_{\boldsymbol{\eta}} \ell^{\star}\left(\boldsymbol{\eta}; \boldsymbol{\varepsilon}, \delta\right)$, where $\ell^{\star}$ refers to the standard likelihood function, $\boldsymbol{\varepsilon}$ is the vector of asset returns, and $\boldsymbol{\eta}$ denotes the entire parameter vector except for δ. Table 1.2 shows the (in-sample) profile likelihood estimates using the EALE. (We show only the 4-component models under study as the results for different numbers of mixture components are

[12] Several models with $k > 4$ were also considered; information criteria never favored them, out-of-sample forecasts were roughly comparable to the $k = 4$ case but increasingly mixed for higher k, and estimations resulted more frequently in estimates with mixture weights close to zero for the additional components. We do not report the results.

[13] By relaxing the equality of the shape parameters constraint, the resulting estimates show an undesired property: While very few mixture components explain the majority of the data with reasonable parameter estimates, the majority of the mixture components maximizes the likelihood based on corner solutions for their shape parameters, which is a classic indicator of overfitting. Using 1000 observations, the MixGED(k, g) and MixStable(k, g) are frequently overfitted, while for larger data sets, the overfitting vanishes, as would be expected, and for the full DJIA sample, MixGED(k, g) estimates with free shape parameters are found to be in line with those reported in Rombouts and Bouaddi (2009).

qualitatively the same.) The optimal choice of δ is between one and two, with a tendency towards one. In addition, Table 1.3 shows that $\delta = 1$ (compared to $\delta = 2$) also improves the out-of-sample density forecasts across models for the majority of data sets in this paper. It also yields lower risk prediction errors in terms of the cumulated root mean squared error over the important VaR level up to 5%.

Our result is also in line with similar studies for simpler GARCH models. In particular, when fitting an asymmetric power GARCH model (APARCH) process with Student's t innovations to 10 national stock indices and the MSCI world index, Brooks et al. (2000) find that the power parameter is significantly different from unity for only one series, whereas it is significantly different from two in nine cases. Similarly, Giot and Laurent (2003) apply a skewed Student's t APARCH model to three national stock markets and three individual stocks. For all but one of the stocks, the estimated δ is rather close to (and statistically indistinguishable from) unity, leading the authors to conclude that "*instead of modeling the conditional variance (GARCH), it is more relevant to model the conditional standard deviation*". More recently, Lejeune (2009) obtains similar results for the S&P 500 and NASDAQ data. Clearly, when working with stable distributions, it no longer makes sense to speak of standard deviations, but rather the scale terms in each of the k components.

Observe that setting $\delta = 1$ implies that the MixStable-GARCH model no longer nests the MixNormal(k, g) model in (1.1) and (1.2), but rather a variant of it, in which the exponent of 2 is replaced by 1.

1.4.2 In-Sample Fit

For assessing in-sample properties, we fit the MixNormal-, MixGED- and MixStable-GARCH models to the seven financial return series under study using the extended likelihood estimator (EALE) from Section 1.3.2. Exemplarily, Table 1.4 shows the parameter estimates of the two MixStable-GARCH models for the DJIA return data.

Table 1.5 shows various in-sample statistics of all models and data sets under study. Out of necessity, the most general model, $\text{A}^1\text{MixStable}(4, 4)$, yields the highest likelihood value for all seven data sets. (This otherwise clear result unfortunately does not carry over to the investigation of out-of-sample forecasts.) As expected, the BIC measure favors less densely parametrized models, with an overall tendency towards the $\text{A}^2\text{MixStable}(2, 2)$ model. We focus on the BIC results as the literature on mixture models provides some theoretical and empirical support for its appropriateness and good performance, in particular for selecting the number of mixture components (see, e.g., Keribin, 2000; Francq et al., 2001; and Frühwirth-Schnatter, 2006, Ch. 4). On the contrary, AIC results are mixed with no clear pattern.

1.4.3 Comparison of Forecasting Performance

While a model as flexible as the MixStable(k, g) should be expected to provide an excellent in-sample fit to virtually any return series compared with more traditional GARCH-type models, the concern remains as to whether the relatively large parametrization, the nontrivial computational aspects of the stable density, and the degeneracy issue associated with mixtures warrant its use. To judge this, we compare the empirical performance of the one-step-ahead predictive cdfs across models using tests for uniformity (see below) as well as concentrating on the left tail, using probability values typical for VaR calculations.

In particular, for a given target probability, λ, typically chosen between 1% and 10%, the VaR delivers an upper bound on losses such that it will be exceeded with probability λ. Conditional on the information given up to time $t-1$, the VaR for period t of one unit of investment is the negative λ-quantile of the conditional return distribution, i.e., $\mathrm{VaR}_{t|\mathcal{F}_{t-1}}(\lambda) = -\inf_x\{x \in \mathbb{R} : \Pr(\varepsilon_t \leq x \mid \mathcal{F}_{t-1}) \geq \lambda\}$ for $0 < \lambda < 1$, where ε_t is the return on an asset or portfolio in period t. In our case, with a continuous, strictly monotone increasing predictive cdf $F_{t|\mathcal{F}_{t-1}}$, we have $\mathrm{VaR}_{t|\mathcal{F}_{t-1}}(\lambda) = -F^{-1}_{t|\mathcal{F}_{t-1}}(\lambda)$. Observe that, while the VaR is considered an inferior risk measure compared to expected shortfall (see, e.g., Dowd, 2005 and the references therein), it is still immensely popular, and the computation of the more sophisticated expected shortfall requires accurate calculation of the VaR.

The choice of the window size is a tuning parameter chosen to maximize the quality of future risk or density forecasts of the particular set of assets under study. It is not necessarily the case that more observations are better, because, with certainty, the proposed model differs in some way from the true data generating process, which itself is likely to not be strictly stationary over long periods of time. A model is possibly, however, a good approximation to reality for short periods of time, though using too short a window results in high variance of the parameters and thus inferior forecasts. To negotiate this bias/variance tradeoff, some experiments with the MixStable-GARCH model and two of the data sets under study indicate that use of a rolling window of sample size 1000 is superior to use of either 500 or 2000. While this value could be optimized further by conducting dedicated experiments per dataset, we use 1000 in all of the following.

For all models considered, we re-estimate the model parameters every 20 trading days (about once a month), so that each estimation contains 2% of new data. Our analysis is based on the realized predictive cdf values obtained from evaluating the one-step-ahead cdf forecasts at the realized returns. If the model is correct, it is well-known that these are uniformly distributed.

Let $\hat{p}_t = \hat{F}_{t|\mathcal{F}_{t-1}}(\varepsilon_t; \hat{\boldsymbol{\theta}}_{t-h})$, $t = 1, \ldots, N$, be the sequence of realized predictive cdf values, noting that, for each t, the parameter vector is estimated using information (in this case, just the past returns) up to and including time $t-h$, where h is a value in $\{1, 2, \ldots, 20\}$, but the entire return series up to time $t-1$ is used in the model filter. Finally, this predictive cdf is evaluated at the actual return at time t. Denote the collection of these N values as vector $\hat{\boldsymbol{p}}$. Further let $\hat{\boldsymbol{p}}^{[s]}$ denote the sorted vector, $\hat{p}_1^{[s]} \leq \hat{p}_2^{[s]} \leq \cdots \leq \hat{p}_N^{[s]}$. The Anderson-Darling (AD) and Cramér-von Mises (CM) test statistics are given respectively by

$$\text{AD} = -N - \sum_{i=1}^{N} \frac{2i-1}{N} \left(\log(\hat{p}_i^{[s]}) + \log(1 - \hat{p}_{N-i+1}^{[s]}) \right)$$

and

$$\text{CM} = \frac{1}{12N} + \sum_{i=1}^{N} \left(\frac{2i-1}{2N} - \hat{p}_i^{[s]} \right)^2 .$$

In addition, we provide test statistics for the Kolmogorov-Smirnov (KS) test for uniformity, as well as the Jarque-Bera (JB) and Shapiro-Wilk (SW) tests for normality after applying the inverse normal cdf transform. We test for serial correlation in $\hat{\boldsymbol{p}}$ (as a proxy for the iid property) and report Ljung-Box (LB) test statistics,

$$\text{LB} = N(N+2) \sum_{i=1}^{m} \frac{\hat{\rho}_i^2}{N-i},$$

where $\hat{\rho}_i$ is the ith autocorrelation from the ith sample autocorrelation function.

Tables 1.6 and 1.7 show the results. It is important to note that here, we are testing the prediction quality over the whole support of the distribution, and not just the left tail (as we do below, for directly testing the quality of value at risk predictions). Except for DAX and NASDAQ, AD and CM are clearly in favor of the stable models with a strong preference for the A^1MixStable-GARCH model. Results for KS and LB are less clear but also pro stable in four out of seven cases. Similar results are obtained for both normality tests, though the JB test appears to favor the A^2MixStable-GARCH model.

We also consider VaR measures dedicated to the left tail, as these are of possibly even greater interest from a risk management perspective. Table 1.8 shows the empirical coverage probabilities (as percentages) for the 1%, 5% and 10% VaR levels along with p-values indicating the severeness of potential risk underestimation. The results for 5% and 10% are mixed, though at the 1% VaR level (arguably the most important VaR level in risk management applications), the A^2MixStable(k, g) model clearly outperforms the other models for most data sets under study.

For further investigations of the VaR prediction quality, we adopt a simple quality measure based on the coverage error over the VaR levels up to $100\lambda\%$, see Kuester et al. (2006). The measure calculates the deviation between predictive cdf and uniform cdf and, thus, captures the excess of percentage violations over the VaR levels, where the deviation is defined as $100(F_U - \hat{F}_e)$ with F_U being the cdf of the standard uniform random variable and $\hat{F}_e$ referring to the empirical cdf formed from $\hat{\boldsymbol{p}}$. Building upon this metric we report the integrated root mean squared error (IRMSE) over the left tail up to the maximal VaR level of interest. The IRMSE employed herein is closely related to the CM statistic but with the sum truncated at $h = \lceil \lambda N \rceil$, i.e.,

$$\mathrm{IRMSE} = \sqrt{\frac{1}{h}\sum_{i=1}^{h}\left(100\frac{2i-1}{2N} - 100\hat{p}_i^{[s]}\right)^2}.$$

The results in Table 1.9 confirm the superiority of the stable models in five out of seven cases at the 1% level, and also at the 10% VaR level.

Finally, we investigate the hit sequence of realized predictive VaR violations,

$$v_t = \mathbb{1}_{\varepsilon_t \leq \hat{q}_t}, \quad \hat{q}_t = \widehat{\mathrm{VaR}}_{t|\mathcal{F}_{t-1}}(\lambda), \tag{1.13}$$

where $\mathbb{1}$ is the indicator function. Under the null of correct conditional coverage, the v_i are iid Bernoulli(λ). From this sequence, the test statistic $\mathrm{LR_{CC}} = \mathrm{LR_{UC}} + \mathrm{LR_{IND}}$ is computed, as proposed in Christoffersen (1998), where $\mathrm{LR_{UC}}$ and $\mathrm{LR_{IND}}$ test for unconditional coverage and independence, respectively. As can be seen from Table 1.10 for the 1% VaR level, the tendency of MixStable(k, g) (in particular of A^{2}MixStable-GARCH) to outperform MixNormal(k, g) and MixGED(k, g) is also corroborated by the Christoffersen test.

1.5 ICA-MixStable-GARCH

A direct generalization of the MixNormal-GARCH model to the multivariate setting with D assets has been investigated by Bauwens et al. (2007) and Haas et al. (2009), the latter model allowing for asymmetries. While of value for a small number of assets, those models will not be practical for even modest portfolios, let alone large ones. In addition, attempting to extend that model to support the multivariate stable distribution is not trivial (but see Lombardi and Veredas, 2009; Bonato, 2011; and the references therein).

A multivariate distribution with MixStable-GARCH marginals can be constructed via use of independent components analysis (ICA); see, e.g., Hyvärinen et al. (2001). Crucially, the resulting multivariate distribution is such that the distribution of a linear combination (as needed to conduct portfolio optimization) is tractable. The method assumes a set of non-Gaussian distributed independent random variables of which linear combinations in the form of time series have been observed. The goal is to recover the original independent random vectors of time series, called the independent components. Once the mixing matrix is known (estimated), the independent components can be modeled and forecasted independently (by, for example, a GARCH-type model). The related concept of conditionally uncorrelated components is discussed in Fan et al. (2008).

An application of the popular iterative FastICA algorithm can be found in Broda and Paolella (2009), where the method is used in a portfolio allocation exercise to estimate the independent components (driven by generalized hyperbolic innovations) of the 30 constituents of the Dow Jones Industrial Average index. To be precise, the ICA variant employed therein maximizes the conditional heteroskedasticity of the independent components. We refer to this specific version as CHICA (Conditionally Heteroskedastic ICA). Unfortunately, the CHICA method has the drawback of requiring finite fourth moments, which is not fulfilled in our setting. However, to the best of our knowledge, the only ICA method dedicated to (time-correlated) stable driven data is found in Fabricius et al. (2001). It is all the more remarkable that in our extensive simulation studies, the method is outperformed by the CHICA method.

Briefly, the CHICA method is a two-step procedure, separating the estimation of the correlation structure from that of the univariate dynamics; details are given in Chen et al. (2006) and Broda and Paolella (2009). A first step is to estimate the expectation of the D assets and obtain the matrix of de-meaned returns $\boldsymbol{Y} \in \mathbb{R}^{D \times T}$. There are several ways of doing this. One could just use the sample mean for each series, though given the fat-tailed nature of the data, this is not efficient. A trimmed mean, or even the median, might be superior, or perhaps best, use of the location term jointly estimated with, say, a MixStable-GARCH model. Whichever is used, further improvement might be realized by (i) estimating it using weighted likelihood, with relatively more weight on more recent observations (see Paolella and Steude, 2008; Broda and Paolella, 2011; and the references therein) and/or (ii) shrinking the D values towards, say, zero.

In the second step, the CHICA method is applied to the residual series $\boldsymbol{Y}$ estimating the mixing matrix $\boldsymbol{A} \in \mathbb{R}^{D \times D}$ that best separates the signals maximizing their GARCH effects. Given the mixing matrix $\boldsymbol{A}$, the D independent components, $\boldsymbol{X} \in \mathbb{R}^{D \times T}$, are obtained by $\boldsymbol{X} = \boldsymbol{A}^{-1}\boldsymbol{Y}$. The third (and only time-consuming) step consists in the estimation and prediction of the D (univariate) independent components based on the MixStable-GARCH model. Observe that the three above steps (in particular, the last one) do not need to be repeated for different portfolio weights. This is crucial if such weights are to be optimized. Note also that each of the D components is endowed with its own, estimated, MixStable-GARCH model, and so each has its own tail index α. This can be contrasted with attempts using the multivariate stable distribution, which has only a single α for all assets.

The distribution of the weighted sum can be computed via standard inversion methods applied to the characteristic function (cf) of the convolution. The stable mixture distribution has a simple expression for its cf, whereas many of the *ad hoc* distributions used in this context, such as GED, do not have such forms, thus precluding their use. (As an aside perhaps worth mentioning, the Student's t distribution is applicable. Let $Z \sim t_\nu$ with zero location and unit scale, and $\nu \in \mathbb{R}_{>0}$. The characteristic function of Z

is

$$\varphi_Z(t;\nu) = \frac{\nu^{\nu/4}|t|^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)} K_{\nu/2}\left(|t|\sqrt{\nu}\right), \tag{1.14}$$

where K_v is the modified Bessel function of the third kind with index v, $v > 0$. Result (1.14) is stated (without reference or derivation) in the reference work of Kotz and Nadarajah (2004, p. 40), while Platen and Heath (2006, p. 37) and Seneta (2004, p. 186) note that it is attributed to Simon R. Hurst, in an unpublished article in 1995 and his dissertation in 1997. Because of the Bessel function, its inversion will be far slower than the stable cf, but it indeed could be used in this context.)

Suppressing the time index for readability, the return R on a portfolio of D assets is a weighted sum of the individual asset returns, and consequently (via the matrix $\mathbf{A}$) also a weighted sum of the independent components X_j with associated weights b_j, say. The distribution of this weighted sum can be computed from its characteristic function (cf), which, from independence, factors into the product of the cfs of the individual components. In our context, each X_j is a k-component stable mixture, with pdf given by $f_{X_j}(x) = \sum_{i=1}^{k} \omega_{j,i} f_S(x; \alpha_{j,i}, \beta_{j,i}, \mu_{j,i}, \sigma_{j,i})$. As the cf of a stable random variable with tail index α, asymmetry parameter β, location μ and scale σ is $\exp\{\imath\mu t - \sigma^\alpha |t|^\alpha (1 - \imath\beta \text{sign}(t) \tan(\alpha\pi/2))\}$, $\imath$ denoting the imaginary unit, the cf of X_j is given by

$$\varphi_{X_j}(t) = \sum_{i=1}^{k} \omega_{j,i} \exp\left\{\imath\mu_{j,i} t - \sigma_{j,i}^{\alpha_{j,i}} |t|^{\alpha_{j,i}} \left(1 - \imath\beta_{j,i}\text{sign}(t)\tan\left(\alpha_{j,i}\frac{\pi}{2}\right)\right)\right\},$$

from which a simple expression for $\varphi_R(t)$ follows. Standard inversion methods can then be applied to φ_R to calculate its pdf or cdf; see, e.g., Paolella (2007, Chapter 1) for details.

For MixStable(k, g), the model lends itself to the computation of portfolio risk measures such as expected shortfall (ES) or VaR due to its simple cf expression. (Interestingly, as shown in Sy (2006), the VaR is a coherent risk measure for portfolios of independent stable Paretian distributed assets, but only if the tail exponent α is the same for all assets.) For continuous return R, the (relative) expected shortfall at $100\lambda\%$ level, $0 < \lambda < 1$, is

$$\text{ES}_R^\lambda = -\mathbb{E}[R|R < q] = -\frac{1}{\lambda}\mathbb{E}[R\mathbb{1}_{R<q}],$$

where $\mathbb{1}$ is the indicator function, and $q = -\text{VaR}_R(\lambda)$ is the $100\lambda\%$ quantile.

If R were just a mixture of stable distributions, one could compute $\mathbb{E}[R\mathbf{1}_{R<q}]$ via the real integral expression for the stable ES in Stoyanov et al. (2006) combined with the general results for the ES of mixtures in Broda and Paolella (2011). However, as R is a convolution of mixtures of stables, this is not applicable. Neither is the general result of Kim et al. (2009a), which requires that the cf be finite in a strip containing the real axis, which is not the case for stable distributions and their mixtures. To compute

$\mathbb{E}[R\mathbf{1}_{R<q}]$, we use the result from Broda (2011) which applies more generally than that given in Kim et al. (2009a),

$$\mathbb{E}[R\mathbf{1}_{R<q}] = \frac{\varphi_R'(0)}{2\imath} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\mathrm{e}^{-\imath tq}\varphi_R'(t)\right]\frac{\mathrm{d}t}{t}, \tag{1.15}$$

where φ_R' is the first derivative of the characteristic function of R, as derived in Appendix A.

Taken together, the computational ingredients are now available to do portfolio optimization in return/ES space, using ICA with the independent components modeled as MixStable processes. The resulting portfolio will, by necessity, have Pareto-like tails, no matter how many assets are involved, unlike in Broda and Paolella (2009), for which, as it uses the normal inverse Gaussian distribution for the components, eventually as D increases, a central limit theorem effect will kick in, and the results will not differ from using a standard Markowitz approach. Given the number of tuning parameters and dependence on the choice of data, the assessment of the performance of the proposed method for portfolio allocation deserves a separate study, and is not pursued here.

1.6 Conclusions and Future Research

An open question in financial econometrics is the suitability of infinite-variance distributions for modeling the unconditional or conditional distribution of asset returns. As discussed at length in the introduction, there is an ongoing debate regarding the maximally existing moment of asset returns. In this paper, we take the stance, as argued by numerous researchers in the stable Paretian field, that the stable Paretian distribution has many positive characteristics for modeling asset returns that outweigh its potential shortcomings. Indeed, we find that a model incorporating mixtures, GARCH, and underlying stable distributions yields a complex but coherent and statistically well motivated model which, based on extensive empirical exercises, delivers both excellent in-sample fit and, most relevantly, admirable out-of-sample forecasting results. In particular, the A^{2}MixStable$(3,2)$, A^{2}MixStable$(3,3)$ and A^{1}MixStable$(4,4)$ models performed best out-of-sample.

Various model extensions suggest themselves for future consideration. There could be value in generalizing the law of motion for the scale terms along the lines of the asymmetric GARCH models in Ding et al. (1993) or Alexander and Lazar (2009). The incorporation of a Markov switching structure, as done in Haas et al. (2004a) and Bauwens et al. (2010), or extension to time-varying component weights, as recently investigated by Bauwens and Storti (2009), could also (and possibly in addition) be entertained.

Finally, as mentioned in Footnote 4, the tempered stable distribution could be considered in place of the stable Paretian, which would address the non-summability issue, and, given its existence of absolute moments, possibly be of value for option pricing; see Poirot and Tankov (2006), Mercuri (2008) and

Kim et al. (2009b) for option pricing with a tempered stable distributional assumption, and Badescu et al. (2008) and Rombouts and Stentoft (2009) for option pricing with the mixed normal distributional assumption.

Appendix

A First Derivative of the Characteristic Function of a Weighted Sum of Mixed Stable Random Variables

For evaluating the ES of a portfolio of mixed stable distributed asset returns based on (1.15), we derive the first derivative of the cf of sum of mixtures of stable Paretian random variables in the general case of asymmetric stable random variables and unrestricted portfolio weights. Similar to the portfolio of D assets in Section 1.5, with summation weights $\boldsymbol{b} \in \mathbb{R}^D$, the portfolio sum is given by $R = \sum_{j=1}^{D} b_j X_j$, where the X_j are independent k-component mixtures of stable random variables,

$$X_j \sim \text{MixStable}(\boldsymbol{\omega}_j, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \boldsymbol{\mu}_j, \boldsymbol{\sigma}_j), \quad j = 1, \ldots, D,$$

with mixture weights $\boldsymbol{\omega}_j \in (0,1)^k$, $\sum_i \omega_{j,i} = 1$, tail indices $\boldsymbol{\alpha}_j \in (0,2]^k$, skewness coefficients $\boldsymbol{\beta}_j \in [-1,1]^k$, location parameters $\boldsymbol{\mu}_j \in \mathbb{R}^k$ and scale terms $\boldsymbol{\sigma}_j \in \mathbb{R}^k_{>0}$. The cf of R is straightforward to calculate and takes the form of

$$\varphi_R\left(t; \boldsymbol{\omega}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\right) = \prod_{j=1}^{D} \sum_{i=1}^{k} \omega_{j,i} \varphi_P\left(t; b_j, \alpha_{j,i}, \beta_{j,i}, \mu_{j,i}, \sigma_{j,i}\right), \tag{1.16}$$

where

$$\varphi_P\left(t; b, \alpha, \beta, \mu, \sigma\right) = \varphi_W\left(t; \alpha, \beta, b\mu, b\sigma\right) = \exp\left\{-\left|\sigma b t\right|^{\alpha}\left(1 - \imath\beta \operatorname{sgn}\left(bt\right) \tan\left(\alpha\frac{\pi}{2}\right)\right) + \imath\mu b t\right\},$$

with $W(\cdot, \alpha, \beta, \mu, \sigma)$ being a stable Paretian random variable with tail index α, asymmetry β, location μ, and scale σ. Straightforward but tedious algebra shows that the first derivative of (1.16) is

$$\varphi_R'\left(t; \boldsymbol{\omega}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\right) = \sum_{i=1}^{D} \exp\left(s_i' + \sum_{j=1, j\neq i}^{D} s_j\right),$$

where

$$s_j = \log \sum_{i=1}^{k} \omega_{j,i} \varphi_P\left(t; b_j, \alpha_{j,i}, \beta_{j,i}, \mu_{j,i}, \sigma_{j,i}\right), \quad j = 1, \ldots, D,$$

$$s_j' = \log \sum_{i=1}^{k} \omega_{j,i} \varphi_P'\left(t; b_j, \alpha_{j,i}, \beta_{j,i}, \mu_{j,i}, \sigma_{j,i}\right), \quad j = 1, \ldots, D,$$

and, for $t \neq 0$,

$$\begin{aligned}\varphi_P'(t;b,\alpha,\beta,\mu,\sigma) &= \frac{\partial}{\partial t}\varphi_P(t;b,\alpha,\beta,\mu,\sigma)\\ &= \varphi_P(t;b,\alpha,\beta,\mu,\sigma)\left(\frac{\alpha}{t}\nu + \imath\mu b\right)\end{aligned}$$

with

$$\nu = -|\sigma b t|^{\alpha}\left(1 - \imath\beta\,\mathrm{sgn}(bt)\tan\left(\alpha\frac{\pi}{2}\right)\right).$$

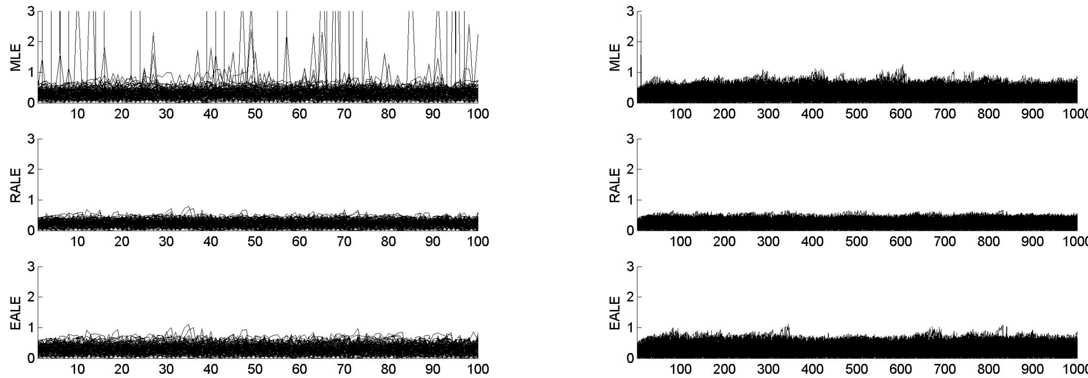
■

B Figures and Tables

Figures and tables are provided on following pages.

MixStable$(2, 2)$

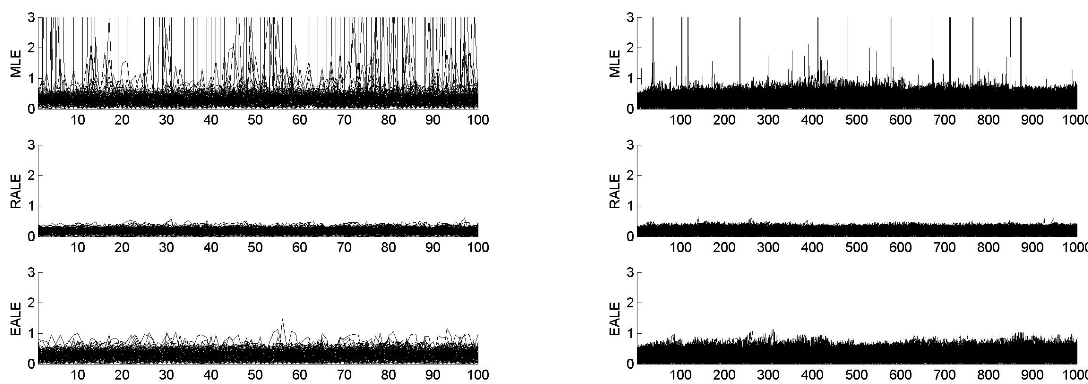
$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
0.315	4e-3	-	0.471	0.046	-	0.785	0.941	-	0.058	0.942	-	-0.721	0.044	-	2	0



sample size 100 sample size 1000

MixStable$(3, 3)$

$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
1.012	4e-13	0.01	0.807	7e-3	0.089	0.648	0.984	0.916	0.019	0.346	0.635	-1.277	0.118	-0.025	2	0



sample size 100 sample size 1000

Figure 1.2: Comparison of MLE, RALE and EALE in presence of local mixture degeneracy. Each panel shows 100 overlaid (maximized) likelihood series (not the augmented likelihood), from which the usual likelihood function is computed, obtained from estimates based on simulated data. Spikes depict either mixture degeneracy (singularities) or almost degenerated local optima. Results are essentially identical for all mixture GARCH models under study. The data generating process roughly corresponds to estimates found in our empirical analysis of DJIA return data.

MixStable(2, 2)

$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
0.004	0.04	-	0.02	0.09	-	0.97	0.91	-	0.07	0.93	-	0.67	-0.05	-	1.96	0

estimator	sample size	P-RMSE arith.	P-RMSE log-geo.		Q-RMSE arith.	Q-RMSE log-geo.	A-RMSE arith.	A-RMSE log-geo.		P-IQR arith.	P-IQR log-geo.	Q-IQR arith.	Q-IQR log-geo.	A-IQR arith.	A-IQR log-geo.
MLE	100	3.56	-60.32	(512)	2.72	-1.70	0.65	-5.98	(84)	3.73	-60.55	1.71	-1.99	0.20	-6.55
RALE		3.70	**-60.20**	(179)	**1.92**	**-2.04**	**0.64**	**-6.00**	(1)	**3.86**	**-60.20**	1.66	**-2.08**	0.23	-6.54
EALE		**3.68**	-60.19	(291)	2.57	-1.80	0.64	-6.00	(106)	4.31	-60.12	**1.65**	-2.03	**0.21**	**-6.56**
MLE	1000	2.47	-60.89	(270)	2.07	-2.11	0.56	-6.03	(17)	2.46	-61.05	0.95	-2.68	0.26	-6.33
RALE		2.93	**-60.57**	(65)	**1.65**	**-2.14**	**0.55**	**-6.05**	(0)	**2.35**	-60.76	**1.04**	**-2.61**	**0.24**	**-6.34**
EALE		**2.67**	-60.65	(106)	1.71	-2.12	0.56	-6.03	(11)	2.55	**-60.78**	1.06	-2.53	0.25	-6.33
MLE	10000	2.66	-61.15	(222)	1.72	-2.24	0.53	-6.07	(14)	2.26	-61.42	0.41	-3.44	0.18	-6.48
RALE		**1.94**	**-61.06**	(25)	**1.67**	**-2.23**	**0.51**	**-6.07**	(0)	**1.83**	-61.28	**0.40**	**-3.48**	0.18	-6.47
EALE		2.36	-61.03	(65)	1.96	-1.91	0.53	-6.06	(1)	1.89	**-61.43**	0.42	-3.43	**0.18**	**-6.47**

MixStable(3, 3)

$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
0.4	0.005	0.07	0.41	0.01	0.09	0.71	0.98	0.92	0.03	0.34	0.63	1.28	0.12	-0.125	1.93	0

estimator	sample size	P-RMSE arith.	P-RMSE log-geo.		Q-RMSE arith.	Q-RMSE log-geo.	A-RMSE arith.	A-RMSE log-geo.		P-IQR arith.	P-IQR log-geo.	Q-IQR arith.	Q-IQR log-geo.	A-IQR arith.	A-IQR log-geo.
MLE	100	7.00	-42.70	(781)	7.02	-0.89	6.55	-2.73	(58)	7.79	-43.82	3.82	-1.42	0.15	-6.99
RALE		6.57	-42.72	(457)	9.06	-1.01	0.54	-5.95	(32)	**6.64**	**-42.68**	**3.54**	-1.44	0.20	-6.83
EALE		**6.55**	**-42.74**	(416)	**8.54**	**-1.04**	**0.53**	**-5.97**	(125)	7.01	-42.67	3.63	**-1.44**	**0.19**	**-6.89**
MLE	1000	5.99	-42.93	(473)	2.66	-1.93	0.46	-5.99	(13)	7.86	-43.15	2.17	-2.12	0.16	-6.76
RALE		5.60	-42.95	(182)	**2.31**	**-1.99**	**0.44**	**-6.02**	(0)	**4.91**	-43.04	2.27	-1.93	0.20	-6.51
EALE		**5.45**	**-42.97**	(91)	2.71	-1.83	0.46	-5.99	(14)	5.04	**-43.11**	**2.04**	**-2.08**	**0.17**	**-6.62**
MLE	10000	4.37	-43.31	(401)	2.17	-2.10	0.44	-6.03	(6)	26.59	-43.69	1.28	-2.57	0.15	-6.73
RALE		4.40	-43.21	(103)	2.42	-1.93	**0.41**	**-6.05**	(0)	3.81	-43.49	1.18	-2.60	0.16	-6.59
EALE		**4.10**	**-43.30**	(59)	**2.16**	**-2.09**	0.44	-6.03	(2)	**3.38**	**-43.90**	**1.14**	**-2.61**	**0.15**	**-6.66**

Table 1.1: Aggregated goodness-of-fit measures based on either parameter errors (P), quantile errors (Q) or acf errors (A) computed from $N = 1000$ estimates based on simulated data. Given values represent error sums, either $\sum_j \xi\left(\epsilon^{\tau}_{j,\cdot}\right)$ (arithmetic) or $1/N \sum_j \log \xi\left(\epsilon^{\tau}_{j,\cdot}\right)$ (log-geometric), where $\xi \in \{\text{RMSE}, \text{IQR}\}$ and $\tau \in \{\text{P}, \text{Q}, \text{A}\}$. As the geometric mean depicts the common factor of the product (being a lower bound to the arithmetic mean), it gives a more robust measure of the aggregated error less affected by outliers than its arithmetic counterpart; without loss of generality we report its log value for reasons of scaling. Values in parentheses denote the number of removed estimates as described in Section 1.3.3. Similar to the case of the P-RMSE, estimates are removed for Q- and A-RMSE where $\sigma_{i,t} \to 0$ or $\sigma_{i,t} \to \infty$. The standard maximum likelihood estimator (gray) is excluded as its likelihood function is ill-defined for mixtures. Entries in boldface denote the smallest error.

data set	model	GARCH power coefficient δ 1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
DAX	MixNormal(4, 4)	-8091.78	-8090.64	-8090.11	-8090.15	-8090.70	-8091.71	**-8093.13**	-8094.89	-8096.94	-8099.25	-8101.68
	MixGED(4, 4)	-8091.78	-8090.64	**-8090.11**	**-8090.15**	**-8090.70**	**-8091.71**	-8093.13	**-8094.89**	-8096.94	-8099.25	-8101.68
	A^1MixStable(4, 4)	**-8091.66**	**-8090.58**	-8090.12	-8090.15	-8090.71	-8091.72	-8093.13	-8094.89	**-8096.93**	**-8099.12**	**-8101.51**
	A^2MixStable(4, 4)	-8103.16	-8102.01	-8101.42	-8101.41	-8101.89	-8102.80	-8104.10	-8105.71	-8107.61	-8109.76	-8112.11
S&P	MixNormal(4, 4)	-6678.03	-6676.27	-6674.94	-6673.98	-6673.37	-6673.07	-6673.06	-6673.31	-6673.80	-6674.51	-6675.42
	MixGED(4, 4)	-6678.03	-6676.27	-6674.94	-6673.98	-6673.37	-6673.07	-6673.06	-6673.31	-6673.80	-6674.51	-6675.42
	A^1MixStable(4, 4)	**-6677.89**	**-6676.13**	**-6674.80**	**-6673.85**	**-6673.24**	**-6672.95**	**-6672.94**	**-6673.20**	**-6673.69**	**-6674.40**	**-6675.32**
	A^2MixStable(4, 4)	-6686.20	-6684.35	-6683.03	-6681.95	-6681.23	-6680.82	-6680.69	-6680.83	-6681.20	-6681.79	-6682.57
DJIA	MixNormal(4, 4)	-6632.67	-6631.47	-6630.62	-6630.10	-6629.87	-6629.92	-6630.21	-6630.73	-6631.45	-6632.35	-6633.42
	MixGED(4, 4)	-6632.72	-6631.50	-6630.62	-6630.10	-6629.87	-6629.92	-6630.21	-6630.73	-6631.45	-6632.35	-6633.42
	A^1MixStable(4, 4)	**-6632.26**	**-6631.08**	**-6630.24**	**-6629.63**	**-6629.33**	**-6629.29**	**-6629.49**	**-6629.93**	**-6630.56**	**-6631.37**	**-6632.36**
	A^2MixStable(4, 4)	-6639.28	-6637.99	-6637.04	-6636.41	-6636.06	-6635.97	-6636.12	-6636.49	-6637.06	-6637.81	-6638.72
NIKKEI	MixNormal(4, 4)	-8459.40	-8458.30	-8457.62	-8457.31	-8457.33	-8457.66	-8458.27	-8459.13	-8460.20	-8461.48	-8462.95
	MixGED(4, 4)	-8459.40	-8458.30	-8457.62	-8457.31	-8457.33	-8457.66	-8458.27	-8459.13	-8460.20	-8461.48	-8462.95
	A^1MixStable(4, 4)	**-8458.99**	**-8457.90**	**-8457.23**	**-8456.93**	**-8456.94**	**-8457.26**	**-8457.85**	**-8458.68**	**-8459.73**	**-8460.97**	**-8462.38**
	A^2MixStable(4, 4)	-8464.85	-8463.98	-8463.55	-8463.51	-8463.80	-8464.39	-8465.20	-8466.26	-8467.51	-8468.95	-8470.54
¥/€	MixNormal(4, 4)	-2627.83	-2626.89	-2626.16	-2625.60	-2625.19	-2624.91	-2624.76	-2624.70	-2624.75	-2624.87	-2625.07
	MixGED(4, 4)	-2627.83	-2626.89	-2626.16	-2625.60	-2625.19	-2624.91	-2624.76	-2624.70	-2624.75	-2624.87	-2625.07
	A^1MixStable(4, 4)	**-2627.83**	**-2626.89**	**-2626.15**	**-2625.59**	**-2625.19**	**-2624.91**	**-2624.76**	**-2624.70**	**-2624.74**	**-2624.87**	**-2625.07**
	A^2MixStable(4, 4)	-2638.27	-2637.16	-2636.22	-2635.49	-2634.92	-2634.51	-2634.22	-2634.05	-2633.98	-2634.01	-2634.12
$/€	MixNormal(4, 4)	-2447.20	-2445.99	-2444.99	-2444.17	-2443.51	-2443.01	-2442.63	-2442.37	-2442.22	-2442.16	-2442.18
	MixGED(4, 4)	-2445.46	-2444.56	-2443.81	-2443.20	-2442.71	-2442.33	-2442.06	-2441.88	-2441.78	-2441.76	-2441.82
	A^1MixStable(4, 4)	**-2443.89**	**-2443.23**	**-2442.74**	**-2442.40**	**-2442.18**	**-2442.07**	**-2441.90**	**-2441.71**	**-2441.59**	**-2441.52**	**-2441.50**
	A^2MixStable(4, 4)	-2447.76	-2447.09	-2446.56	-2446.15	-2445.84	-2445.65	-2445.42	-2445.27	-2445.18	-2445.17	-2445.22
NASDAQ	MixNormal(4, 4)	-8138.21	-8137.30	-8136.79	-8136.62	-8136.77	-8137.19	-8137.86	-8138.75	-8139.85	-8141.13	-8142.57
	MixGED(4, 4)	-8138.21	-8137.30	-8136.79	-8136.62	-8136.77	-8137.19	-8137.86	-8138.75	-8139.85	-8141.13	-8142.57
	A^1MixStable(4, 4)	**-8134.43**	**-8133.49**	**-8132.96**	**-8132.78**	**-8132.90**	**-8133.30**	**-8133.95**	**-8134.82**	**-8135.88**	**-8137.12**	**-8138.53**
	A^2MixStable(4, 4)	-8139.44	-8138.49	-8137.92	-8137.71	-8137.80	-8138.17	-8138.79	-8139.63	-8140.67	-8141.89	-8143.25

Table 1.2: Profile likelihood results for different GARCH power parameters δ and all 4-component models and data sets under study. Each row corresponds to fitting the model to the entire return series of 20 years (equity) or 10 years (FX), respectively. Entries in boldface denote the best (in-sample) fits. Underlined numbers denote the best outcome per model. Gray values denote MixStable-GARCH estimates for which the power restriction $\delta < \alpha$ (necessary for $\mathbb{E}[|\varepsilon_t|^\delta] < \infty$) is violated.

model	δ	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Anderson-Darling								
MixNormal(2, 2)	1	0.86	1.00	1.27	**0.93**	**1.54**	**1.52**	**1.32**
	2	**0.71**	**0.89**	**1.19**	1.10	1.72	2.01*	1.37
MixNormal(3, 2)	1	0.65	**0.54**	**0.59**	**0.73**	**1.49**	**1.33**	**0.43**
	2	**0.57**	0.67	0.59	0.76	1.53	1.92	0.49
MixNormal(3, 3)	1	0.55	0.53	**0.54**	**0.60**	**1.41**	**1.51**	**0.42**
	2	**0.52**	**0.50**	0.63	0.70	1.47	1.99*	0.43
MixNormal(4, 4)	1	0.69	**0.37**	**0.42**	**0.54**	**1.22**	**1.59**	**0.47**
	2	**0.58**	0.42	0.66	0.63	1.29	1.61	0.48
MixGED(2, 2)	1	0.72	**0.63**	**0.67**	**0.73**	**1.44**	**1.41**	**0.30**
	2	**0.64**	0.71	0.68	0.85	1.72	1.83	0.59
MixGED(3, 2)	1	0.76	0.56	0.52	**0.62**	**1.45**	**1.09**	0.52
	2	**0.73**	**0.53**	**0.46**	0.73	1.59	1.81	**0.33**
MixGED(3, 3)	1	0.68	**0.51**	**0.49**	**0.48**	1.45	**1.30**	0.46
	2	**0.56**	0.51	0.82	0.71	**1.45**	1.87	**0.38**
MixGED(4, 4)	1	0.72	**0.41**	**0.59**	**0.45**	**1.21**	**1.47**	0.51
	2	**0.58**	0.56	0.82	0.65	1.23	1.79	**0.46**
Cramér-von Mises								
MixNormal(2, 2)	1	0.12	0.12	0.22	**0.16**	**0.30**	**0.30**	0.23
	2	**0.11**	**0.11**	**0.20**	0.18	0.34	0.41*	**0.19**
MixNormal(3, 2)	1	**0.06**	**0.07**	0.07	0.12	**0.27**	**0.25**	**0.04**
	2	0.07	0.08	**0.06**	**0.11**	0.28	0.39*	0.06
MixNormal(3, 3)	1	0.06	0.08	**0.06**	0.08	**0.25**	**0.28**	**0.04**
	2	**0.05**	**0.06**	0.07	**0.07**	0.28	0.39*	0.04
MixNormal(4, 4)	1	0.07	**0.06**	**0.05**	0.08	**0.22**	**0.29**	0.04
	2	**0.06**	0.06	0.07	**0.07**	0.23	0.30	**0.04**
MixGED(2, 2)	1	**0.09**	**0.08**	**0.09**	**0.10**	**0.27**	**0.26**	**0.03**
	2	0.09	0.08	0.10	0.13	0.34	0.36*	0.06
MixGED(3, 2)	1	**0.08**	0.07	0.06	**0.09**	**0.26**	**0.20**	0.05
	2	0.09	**0.06**	**0.05**	0.10	0.29	0.36*	**0.03**
MixGED(3, 3)	1	0.07	0.07	**0.06**	**0.06**	**0.26**	**0.24**	0.04
	2	**0.06**	**0.06**	0.09	0.07	0.27	0.36*	**0.04**
MixGED(4, 4)	1	0.07	**0.05**	**0.07**	**0.06**	0.22	**0.27**	0.05
	2	**0.06**	0.07	0.09	0.06	**0.22**	0.33	**0.04**
Integrated root mean squared error, IRMSE, up to the 5% VaR level								
MixNormal(2, 2)	1	**0.19**	0.36	**0.22**	**0.31**	**0.34**	0.53	**0.20**
	2	0.21	**0.32**	0.29	0.39	0.36	**0.51**	0.25
MixNormal(3, 2)	1	**0.17**	**0.25**	**0.16**	**0.28**	**0.47**	0.69	0.49
	2	0.22	0.27	0.16	0.35	0.51	**0.49**	**0.35**
MixNormal(3, 3)	1	**0.10**	0.23	**0.10**	**0.28**	0.47	0.62	0.53
	2	0.22	**0.22**	0.14	0.34	**0.46**	**0.37**	**0.37**
MixNormal(4, 4)	1	**0.20**	0.22	**0.10**	**0.27**	**0.42**	0.59	0.64
	2	0.24	**0.19**	0.11	0.37	0.48	**0.40**	**0.51**
MixGED(2, 2)	1	**0.20**	**0.27**	**0.16**	**0.31**	0.40	**0.31**	0.35
	2	0.23	0.28	0.25	0.34	**0.36**	0.34	**0.23**
MixGED(3, 2)	1	**0.19**	0.29	**0.12**	**0.26**	**0.46**	0.54	0.60
	2	0.25	**0.23**	0.13	0.35	0.50	**0.34**	**0.32**
MixGED(3, 3)	1	**0.16**	**0.23**	**0.12**	**0.28**	0.47	**0.33**	0.61
	2	0.24	0.26	0.15	0.34	**0.44**	0.39	**0.39**
MixGED(4, 4)	1	**0.17**	**0.22**	**0.10**	**0.27**	**0.44**	**0.37**	0.64
	2	0.25	0.24	0.11	0.38	0.47	0.45	**0.54**

Table 1.3: Anderson-Darling and Cramér-von Mises test statistics as well as IRMSE values for all MixNormal- and MixGED-GARCH models and data sets under study. Entries in boldface denote the best outcomes. For AD and CM, ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. All values are based on evaluating the one-step-ahead out-of-sample distribution forecasts at the observed return data, estimated throughout a rolling window of $1,000$ trading days (updated every 20 days), covering 10 years of equity returns (July 7, 1999 to July 7, 2009) and 5 years (July 7, 2004 to July 7, 2009) of FX returns.

	A^1Mix-Stable $(2,2)$	A^1Mix-Stable $(3,2)$	A^1Mix-Stable $(3,3)$	A^1Mix-Stable $(4,4)$	A^2Mix-Stable $(2,2)$	A^2Mix-Stable $(3,2)$	A^2Mix-Stable $(3,3)$	A^2Mix-Stable $(4,4)$
$\gamma_{0,1}$	4.8e-13	1.6e-20	0.130	1.2e-20	2.1e-18	3.4e-13	7.1e-14	0.003
	(1.3e-15)	(9.5e-27)	(0.049)	(1.4e-27)	(9.9e-24)	(5.7e-16)	(9.6e-17)	(0.001)
$\gamma_{0,2}$	0.012	0.012	8.9e-25	0.010	0.007	0.007	0.008	2.7e-20
	(0.004)	(0.003)	(1.8e-39)	(0.003)	(0.002)	(0.002)	(0.003)	(8.6e-27)
$\gamma_{0,3}$		2.682	0.005	0.438		0.964	0.002	0.002
		(0.387)	(0.002)	(0.143)		(3.028)	(0.001)	(0.001)
$\gamma_{0,4}$				6.0e-13				0.042
				(5.2e-16)				(0.016)
$\gamma_{1,1}$	0.015	0.014	0.339	0.014	0.010	0.010	0.011	0.119
	(0.004)	(0.003)	(0.090)	(0.004)	(0.003)	(0.002)	(0.003)	(0.018)
$\gamma_{1,2}$	0.112	0.105	0.011	0.123	0.074	0.074	0.118	0.010
	(0.013)	(0.013)	(0.003)	(0.018)	(0.007)	(0.008)	(0.016)	(0.003)
$\gamma_{1,3}$			0.067	0.442			0.021	0.021
			(0.008)	(0.140)			(0.005)	(0.005)
$\gamma_{1,4}$				0.023				0.101
				(0.005)				(0.042)
ψ_1	0.984	0.984	0.778	0.970	0.981	0.981	0.976	0.911
	(0.004)	(0.003)	(0.057)	(0.008)	(0.006)	(0.005)	(0.007)	(0.013)
ψ_2	0.911	0.917	0.978	0.904	0.936	0.936	0.908	0.976
	(0.010)	(0.009)	(0.006)	(0.013)	(0.006)	(0.006)	(0.011)	(0.008)
ψ_3			0.943	0.663			0.979	0.979
			(0.007)	(0.080)			(0.005)	(0.005)
ψ_4				0.980				0.893
				(0.004)				(0.033)
ω_1	0.439	0.420	0.046	0.101	0.145	0.145	0.089	0.438
	(0.024)	(0.057)	(0.012)	(0.024)	(0.032)	(0.037)	(0.020)	(0.065)
ω_2	0.561	0.571	0.131	0.477	0.855	0.855	0.498	0.086
	(red.)	(0.007)	(0.039)	(0.061)	(red.)	(4.1e-6)	(0.066)	(0.033)
ω_3		0.009	0.823	0.019		1.6e-10	0.413	0.408
		(red.)	(red.)	(0.011)		(red.)	(red.)	(0.107)
ω_4				0.403				0.068
				(red.)				(red.)
μ_1	0.138	0.118	0.846	0.048				
	(0.033)	(0.026)	(0.165)	(0.039)				
μ_2	0.108	0.054	0.081	0.129				
	(red.)	(0.030)	(0.037)	(0.055)				
μ_3		2.028	0.035	1.226				
		(red.)	(red.)	(0.195)				
μ_4				0.197				
				(red.)				
α	1.917	2.000	1.989	1.997	1.890	1.890	1.892	1.894
	(0.025)	(0.004)	(4.7e-5)	(0.010)	(0.013)	(0.013)	(0.013)	(0.013)
β					0.707	0.706	0.744	0.781
					(0.061)	(0.065)	(0.027)	(0.059)
LL	-6652.061	-6641.523	-6639.499	-6632.264	-6644.757	-6644.756	-6639.442	-6639.277

Table 1.4: Maximum likelihood parameter estimates for the MixStable-GARCH models under study based on the entire set of DJIA return data (20 years). Standard errors are given in parentheses. The shortcut "red." stands for "redundant" and indicates that the parameter is not estimated but recovered from the corresponding imposed constraints.

model	free param.	crit.	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
MixNormal(2, 2)	8	LL	-8120.9	-6716.4	-6660.3	-8476.0	-2634.8	-2455.1	-8163.5
		AIC	16257.8	13448.8	13336.6	16968.0	**5285.5**	4926.1	16343.1
		BIC	16310.0	13501.0	13388.7	17020.0	**5332.2**	4972.8	16395.4
MixNormal(3, 2)	11	LL	-8097.2	-6688.3	-6641.5	-8466.7	-2632.7	-2450.2	-8151.8
		AIC	**16216.3**	13398.6	13305.0	16955.4	5287.3	4922.4	16325.5
		BIC	16288.1	13470.4	13376.8	17027.0	5351.5	4986.6	16397.4
MixNormal(3, 3)	13	LL	-8097.1	-6686.1	-6639.3	-8465.2	-2629.8	-2449.6	-8146.1
		AIC	16220.3	13398.3	13304.7	16956.4	5285.6	4925.3	16318.2
		BIC	16305.1	13483.1	13389.5	17041.0	5361.4	5001.1	16403.1
MixNormal(4, 4)	18	LL	-8091.8	-6678.0	-6632.7	-8459.4	-2627.8	-2447.2	-8138.2
		AIC	16219.6	**13392.1**	13301.3	16954.8	5291.7	4930.4	16312.4
		BIC	16337.0	13509.5	13418.7	17071.8	5396.6	5035.4	16430.0
MixGED(2, 2)	9	LL	-8116.2	-6693.3	-6644.9	-8471.1	-2634.8	-2446.3	-8158.6
		AIC	16250.4	13404.6	13307.9	16960.1	5287.5	**4910.5**	16335.2
		BIC	16309.2	13463.3	13366.6	**17018.6**	5340.0	**4963.1**	16394.0
MixGED(3, 2)	12	LL	-8097.0	-6686.7	-6637.6	-8465.1	-2632.6	-2445.7	-8151.1
		AIC	16218.0	13397.4	**13299.2**	**16954.3**	5289.3	4915.5	16326.2
		BIC	16296.3	13475.7	13377.5	17032.3	5359.2	4985.5	16404.6
MixGED(3, 3)	14	LL	-8097.0	-6685.5	-6636.7	-8464.4	-2629.8	-2445.5	-8146.1
		AIC	16222.0	13399.1	13301.4	16956.7	5287.6	4918.9	16320.2
		BIC	16313.3	13490.4	13392.7	17047.7	5369.2	5000.6	16411.6
MixGED(4, 4)	19	LL	-8091.8	-6678.0	-6632.7	-8459.4	-2627.8	-2445.5	-8138.2
		AIC	16221.6	13394.1	13303.4	16956.8	5293.7	4928.9	16314.4
		BIC	16345.5	13518.0	13427.4	17080.3	5404.4	5039.8	16438.6
A^1MixStable(2, 2)	9	LL	-8103.0	-6699.9	-6652.1	-8471.2	-2634.4	-2450.1	-8159.8
		AIC	16224.0	13417.8	13322.1	16960.4	5286.8	4918.1	16337.7
		BIC	**16282.7**	13476.6	13380.8	17018.9	5339.3	4970.6	16396.5
A^1MixStable(3, 2)	12	LL	-8097.2	-6688.3	-6641.5	-8466.7	-2632.7	-2447.7	-8137.4
		AIC	16218.3	13400.7	13307.0	16957.4	5289.3	4919.3	**16298.8**
		BIC	16296.6	13478.9	13385.3	17035.5	5359.3	4989.4	16377.3
A^1MixStable(3, 3)	14	LL	-8096.5	-6686.1	-6639.5	-8463.8	-2629.8	-2446.3	-8142.2
		AIC	16221.0	13400.3	13307.0	16955.6	5287.6	4920.7	16312.4
		BIC	16312.4	13491.6	13398.3	17046.7	5369.2	5002.4	16403.9
A^1MixStable(4, 4)	19	LL	**-8091.7**	**-6677.9**	**-6632.3**	**-8459.0**	**-2627.8**	**-2443.9**	**-8134.4**
		AIC	16221.3	13393.8	13302.5	16956.0	5293.7	4925.8	16306.9
		BIC	16345.3	13517.7	13426.5	17079.5	5404.4	5036.6	16431.0
A^2MixStable(2, 2)	9	LL	-8103.3	-6692.0	-6644.8	-8475.0	-2638.5	-2450.7	-8146.9
		AIC	16224.6	13402.0	13307.5	16968.0	5295.1	4919.5	16311.7
		BIC	16283.3	**13460.7**	**13366.2**	17026.5	5347.5	4972.0	**16370.6**
A^2MixStable(3, 2)	12	LL	-8103.3	-6692.0	-6644.8	-8471.0	-2638.5	-2450.0	-8147.7
		AIC	16228.6	13406.0	13311.5	16964.1	5299.1	4921.9	16317.4
		BIC	16300.4	13477.8	13383.3	17035.6	5363.2	4986.1	16389.3
A^2MixStable(3, 3)	13	LL	-8103.2	-6686.3	-6639.4	-8468.1	-2638.4	-2448.8	-8141.1
		AIC	16232.3	13398.5	13304.9	16962.2	5302.9	4923.6	16308.2
		BIC	16317.2	13483.3	13389.7	17046.7	5378.7	4999.4	16393.2
A^2MixStable(4, 4)	17	LL	-8103.2	-6686.2	-6639.3	-8464.9	-2638.3	-2447.8	-8139.4
		AIC	16240.3	13406.4	13312.6	16963.7	5310.5	4929.5	16312.9
		BIC	16351.3	13517.3	13423.4	17074.2	5409.6	5028.7	16424.0

Table 1.5: Estimation results for all multi-component mixture GARCH models using $\delta = 1$ and all data sets under study (20 years of equity returns, 10 years of FX returns). The three-row blocks show the log-likelihood value (LL) and the corresponding information criteria (AIC and BIC). Entries in boldface denote the best column-wise (in-sample) fits.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Anderson-Darling							
MixNormal(2, 2)	0.86	1.00	1.27	0.93	1.54	1.52	1.32
MixNormal(3, 2)	0.65	0.54	0.59	0.73	1.49	1.33	0.43
MixNormal(3, 3)	**0.55**	0.53	0.54	0.60	1.41	1.51	0.42
MixNormal(4, 4)	0.69	0.37	0.42	0.54	1.22	1.59	0.47
MixGED(2, 2)	0.72	0.63	0.67	0.73	1.44	1.41	**0.30**
MixGED(3, 2)	0.76	0.56	0.52	0.62	1.45	1.09	0.52
MixGED(3, 3)	0.68	0.51	0.49	0.48	1.45	1.30	0.46
MixGED(4, 4)	0.72	0.41	0.59	0.45	1.21	1.47	0.51
A^1MixStable(2, 2)	0.73	0.49	0.55	0.92	1.36	1.20	0.35
A^1MixStable(3, 2)	0.70	0.45	0.43	0.65	1.48	**1.01**	0.31
A^1MixStable(3, 3)	0.66	0.43	0.36	0.44	1.43	1.38	0.36
A^1MixStable(4, 4)	0.63	**0.31**	**0.28**	**0.34**	1.24	1.48	0.44
A^2MixStable(2, 2)	1.12	1.05	0.67	0.59	**0.84**	1.88	1.61
A^2MixStable(3, 2)	1.16	1.02	0.58	0.58	0.85	1.71	1.59
A^2MixStable(3, 3)	1.22	0.95	0.56	0.58	0.85	1.68	1.59
A^2MixStable(4, 4)	1.23	0.99	0.56	0.59	0.86	1.70	1.63
Cramér-von Mises							
MixNormal(2, 2)	0.12	0.12	0.22	0.16	0.30	0.30	0.23
MixNormal(3, 2)	0.06	0.07	0.07	0.12	0.27	0.25	0.04
MixNormal(3, 3)	**0.06**	0.08	0.06	0.08	0.25	0.28	0.04
MixNormal(4, 4)	0.07	0.06	0.05	0.08	0.22	0.29	0.04
MixGED(2, 2)	0.09	0.08	0.09	0.10	0.27	0.26	**0.03**
MixGED(3, 2)	0.08	0.07	0.06	0.09	0.26	0.20	0.05
MixGED(3, 3)	0.07	0.07	0.06	0.06	0.26	0.24	0.04
MixGED(4, 4)	0.07	0.05	0.07	0.06	0.22	0.27	0.05
A^1MixStable(2, 2)	0.09	0.07	0.08	0.15	0.25	0.23	0.04
A^1MixStable(3, 2)	0.08	0.05	0.05	0.10	0.27	**0.19**	0.03
A^1MixStable(3, 3)	0.06	0.06	0.04	0.05	0.26	0.24	0.04
A^1MixStable(4, 4)	0.06	**0.05**	**0.03**	**0.04**	0.22	0.26	0.04
A^2MixStable(2, 2)	0.14	0.15	0.10	0.07	**0.15**	0.41*	0.27
A^2MixStable(3, 2)	0.14	0.14	0.08	0.06	0.16	0.37*	0.26
A^2MixStable(3, 3)	0.15	0.13	0.07	0.06	0.16	0.36*	0.27
A^2MixStable(4, 4)	0.15	0.14	0.07	0.06	0.16	0.37*	0.27
Kolmogorov-Smirnov (test statistics are scaled up by factor 100)							
MixNormal(2, 2)	1.48	1.44	1.97	1.73	3.23	3.76*	2.41
MixNormal(3, 2)	1.28	1.64	1.36	1.47	3.39	3.47	1.07
MixNormal(3, 3)	**1.19**	1.57	1.25	1.25	3.40	3.58	0.96*
MixNormal(4, 4)	1.38	1.32	1.17	1.34	3.12	3.71	1.19
MixGED(2, 2)	1.33	1.62	1.33	1.44	3.13	3.63	1.09
MixGED(3, 2)	1.36	1.47	1.21	1.26	3.39	**2.77**	1.05
MixGED(3, 3)	1.33	1.50	1.36	1.15	3.40	3.33	0.97*
MixGED(4, 4)	1.42	**1.06**	1.34	1.03	3.19	3.37	1.06
A^1MixStable(2, 2)	1.29	1.25	1.31	1.73	3.11	3.14	1.11
A^1MixStable(3, 2)	1.27	1.54	1.21	1.59	3.32	2.79	**0.86****
A^1MixStable(3, 3)	1.33	1.48	1.21	1.10	3.40	3.33	1.04
A^1MixStable(4, 4)	1.28	1.11	**1.02**	**0.90****	3.25	3.35	1.14
A^2MixStable(2, 2)	2.17	1.63	1.74	1.54	**2.86**	4.27**	2.18
A^2MixStable(3, 2)	2.21	1.79	1.30	1.41	2.87	3.98*	2.03
A^2MixStable(3, 3)	2.21	1.67	1.26	1.46	2.89	3.88*	2.16
A^2MixStable(4, 4)	2.20	1.77	1.32	1.51	2.89	3.88*	2.16

Table 1.6: Anderson-Darling, Cramér-von Mises and Kolmogorov-Smirnov test statistics for all models (with $\delta = 1$) and data sets under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 1.3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Ljung-Box ($m = 20$ lags)							
MixNormal(2, 2)	26.44	32.97*	30.66	16.95	17.46	18.57	31.05
MixNormal(3, 2)	26.74	33.15*	30.46	17.01	17.59	18.55	31.62*
MixNormal(3, 3)	26.63	33.26*	30.68	17.49	17.65	18.50	30.92
MixNormal(4, 4)	**26.25**	33.77*	30.78	17.95	17.99	18.03	31.86*
MixGED(2, 2)	26.62	33.61*	31.12	**16.94**	17.46	**17.55**	30.92
MixGED(3, 2)	26.73	33.15*	30.69	17.12	17.78	17.95	31.83*
MixGED(3, 3)	26.90	33.41*	31.33	17.96	17.64	17.96	31.37
MixGED(4, 4)	26.51	33.61*	30.62	18.10	17.82	18.20	31.46*
A^1MixStable(2, 2)	26.46	**32.87****	**29.83**	16.96	17.34	18.49	**30.49**
A^1MixStable(3, 2)	26.29	33.35*	30.63	19.64	17.67	19.17	31.50*
A^1MixStable(3, 3)	26.33	33.55*	30.72	17.59	17.77	17.94	30.78
A^1MixStable(4, 4)	26.44	33.94*	31.25	18.24	17.67	18.29	31.45*
A^2MixStable(2, 2)	27.21	33.45*	30.00	17.83	16.17	17.66	31.30
A^2MixStable(3, 2)	26.95	33.12*	29.86	17.92	**16.14**	17.74	31.58*
A^2MixStable(3, 3)	27.00	33.14*	29.99	17.89	16.22	17.94	31.55*
A^2MixStable(4, 4)	27.04	33.14*	29.85	17.91	16.21	17.86	31.42*
Jarque-Bera							
MixNormal(2, 2)	28.31***	51.49***	18.88***	**2.11**	9.73**	9.71**	80.46***
MixNormal(3, 2)	17.25***	3.01	5.69	3.19	14.94***	17.53***	7.01*
MixNormal(3, 3)	19.61***	198.32***	52.25***	4.95	22.28***	16.10***	65.26***
MixNormal(4, 4)	4.61	70.44***	5.57	5.42	13.52***	20.68***	56.53***
MixGED(2, 2)	2.24	9.63**	9.63**	3.14	9.82**	6.64*	19.23***
MixGED(3, 2)	15.18***	9.35**	9.94**	3.43	14.12***	3.79	7.23*
MixGED(3, 3)	17.54***	12.13***	4.45	4.22	19.36***	5.50	18.01***
MixGED(4, 4)	4.97	9.10**	9.06**	4.69	15.94***	6.67*	16.18***
A^1MixStable(2, 2)	2.06	1.28	8.40**	2.23	10.47**	5.49	0.80
A^1MixStable(3, 2)	2.04	1.63	2.72	3.31	14.52***	4.14	1.28
A^1MixStable(3, 3)	1.31	2.47	2.22	5.74	18.98***	4.88	1.71
A^1MixStable(4, 4)	1.17	0.79	1.45	3.51	10.79***	7.91**	2.23
A^2MixStable(2, 2)	**0.28**	0.06*	0.54	2.75	3.70	8.56**	0.38
A^2MixStable(3, 2)	0.32	**0.05****	**0.39**	2.83	3.41	6.10*	1.77
A^2MixStable(3, 3)	0.67	0.23	0.42	3.31	3.78	**3.13**	**0.03****
A^2MixStable(4, 4)	0.69	0.25	0.51	3.11	**3.41**	3.25	0.17
Shapiro-Wilk (test statistic ν is transformed by $1000(1-\nu)$)							
MixNormal(2, 2)	3.16***	3.87***	2.65***	**1.23**	3.11**	2.01	6.04***
MixNormal(3, 2)	2.32***	0.94	1.68**	1.32*	4.00***	2.91**	1.63**
MixNormal(3, 3)	2.40***	6.94***	4.53***	1.65**	4.72***	2.69*	5.37***
MixNormal(4, 4)	1.16	4.17***	1.44**	1.60**	3.37**	3.13**	5.43***
MixGED(2, 2)	0.92	1.78**	2.23***	1.25	3.10**	1.74	3.05***
MixGED(3, 2)	2.27***	1.72**	2.30***	1.27*	3.87***	**1.01**	2.22***
MixGED(3, 3)	2.25***	1.93***	1.30*	1.52**	4.49***	1.49	3.68***
MixGED(4, 4)	1.19	1.50**	2.24***	1.50**	3.87***	1.55	3.62***
A^1MixStable(2, 2)	0.67	0.58	1.78**	1.31*	3.25**	1.32	0.91
A^1MixStable(3, 2)	0.57	0.55	1.00	1.27*	3.93***	1.17	**0.82**
A^1MixStable(3, 3)	0.57	0.59	0.90	1.56**	4.43***	1.35	1.23
A^1MixStable(4, 4)	**0.56**	**0.41**	0.81	1.34*	3.12**	1.86	1.43**
A^2MixStable(2, 2)	1.51**	1.16	0.75	1.26*	**2.06**	2.19	1.76**
A^2MixStable(3, 2)	1.55**	1.15	**0.74**	1.33*	2.14	1.85	1.73**
A^2MixStable(3, 3)	1.43**	1.15	0.77	1.27*	2.11	1.34	1.75**
A^2MixStable(4, 4)	1.44**	1.08	0.76	1.28*	2.14	1.36	1.85***

Table 1.7: Ljung-Box, Jarque-Bera and Shapiro-Wilk test statistics for all models (with $\delta = 1$) and data sets under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 1.3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
MixNormal(2, 2)	1.01	1.11	1.21	0.87	1.60**	1.15	**0.96**
MixNormal(3, 2)	0.96	1.08	1.06	0.87	1.68**	1.11	0.60
MixNormal(3, 3)	0.84	1.09	1.13	0.84	1.63**	1.20	0.57
MixNormal(4, 4)	0.92	1.12	1.03	0.89	1.62**	1.19	0.61
MixGED(2, 2)	**1.00**	1.11	1.11	0.90	1.75**	1.40	0.81
MixGED(3, 2)	0.96	1.18	1.07	0.87	1.67**	1.11	0.63
MixGED(3, 3)	0.88	1.10	1.03	0.80	1.76**	1.32	0.59
MixGED(4, 4)	0.88	1.10	1.03	0.85	1.72**	1.20	0.64
A^1MixStable(2, 2)	1.11	1.13	1.11	0.87	1.75**	1.16	0.68
A^1MixStable(3, 2)	1.09	1.17	1.03	0.88	1.60**	1.13	0.62
A^1MixStable(3, 3)	1.02	1.14	1.17	0.85	1.63**	1.22	0.70
A^1MixStable(4, 4)	0.88	1.06	**1.02**	0.92	1.63**	1.20	0.67
A^2MixStable(2, 2)	0.77	**0.98**	1.05	0.90	1.46	1.13	0.71
A^2MixStable(3, 2)	0.77	0.98	1.07	0.91	**1.40**	1.09	0.68
A^2MixStable(3, 3)	0.77	0.98	1.05	**0.98**	1.42	**1.08**	0.64
A^2MixStable(4, 4)	0.78	0.98	1.04	0.93	1.46	1.09	0.66
5% VaR							
MixNormal(2, 2)	5.42	5.48	5.46	5.83**	5.02	4.28	5.49
MixNormal(3, 2)	5.38	5.35	5.22	5.70*	5.15	3.96	4.81
MixNormal(3, 3)	5.23	5.20	5.22	5.58	5.17	4.01	4.60
MixNormal(4, 4)	5.33	**5.04**	5.19	5.56	5.18	3.96	4.45
MixGED(2, 2)	5.30	5.20	5.45	5.90**	**5.02**	4.51	5.09
MixGED(3, 2)	**5.22**	5.34	5.30	5.77**	5.33	4.15	4.61
MixGED(3, 3)	5.26	5.13	5.35	5.80**	5.23	4.44	4.58
MixGED(4, 4)	5.33	5.22	5.30	5.52	5.20	4.24	4.49
A^1MixStable(2, 2)	5.46	5.33	5.15	5.86**	5.10	4.06	**5.04**
A^1MixStable(3, 2)	5.27	5.21	5.05	5.49	5.15	3.95	4.83
A^1MixStable(3, 3)	5.31	5.14	5.05	5.63*	5.15	4.20	4.37
A^1MixStable(4, 4)	5.29	4.94	**5.03**	**5.48**	5.20	4.19	4.26
A^2MixStable(2, 2)	5.47	5.47	5.36	5.67*	5.55	**4.61**	5.61*
A^2MixStable(3, 2)	5.46	5.46	5.32	5.67*	5.55	4.61	5.40
A^2MixStable(3, 3)	5.51	5.40	5.28	5.60*	5.65	4.54	5.37
A^2MixStable(4, 4)	5.51	5.43	5.26	5.64*	5.64	4.53	5.52
10% VaR							
MixNormal(2, 2)	10.69	11.21**	10.60	10.62	9.68	9.07	11.17**
MixNormal(3, 2)	10.75	10.99*	10.97*	10.55	9.32	8.88	10.24
MixNormal(3, 3)	**10.56**	10.71	10.84*	10.64	9.10	8.72	10.16
MixNormal(4, 4)	10.76	**10.63**	10.85*	10.47	9.07	8.79	9.69
MixGED(2, 2)	10.98*	11.00**	10.63	10.61	9.57	8.98	**10.05**
MixGED(3, 2)	10.87*	10.94*	10.90*	10.63	9.36	8.82	9.74
MixGED(3, 3)	10.63	10.76	10.99*	10.53	9.10	8.88	9.70
MixGED(4, 4)	10.73	10.72	10.79	10.53	9.12	8.88	9.67
A^1MixStable(2, 2)	10.80	10.83*	**10.40**	10.53	9.55	9.07	10.42
A^1MixStable(3, 2)	10.68	10.88*	10.85*	**10.24**	9.40	9.04	10.25
A^1MixStable(3, 3)	10.66	10.74	10.68	10.42	9.10	8.59	10.20
A^1MixStable(4, 4)	10.60	10.64	10.69	10.37	9.21	8.59	9.71
A^2MixStable(2, 2)	11.59***	11.26**	10.73	10.99*	9.61	9.15	11.26**
A^2MixStable(3, 2)	11.69***	11.36**	10.77	10.95*	9.61	9.29	11.22**
A^2MixStable(3, 3)	11.74***	11.43***	10.71	10.95*	9.71	**9.29**	11.28**
A^2MixStable(4, 4)	11.77***	11.36**	10.66	10.96*	**9.73**	9.29	11.34**

Table 1.8: Predicted VaR coverage percentages (point estimates) at the 1%, 5% and 10% level for all models under study (with $\delta = 1$). Entries in boldface denote the best (closest to the true value) estimate. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 1.3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
MixNormal(2, 2)	0.10	0.22	0.12	0.11	0.32	0.10	**0.06**
MixNormal(3, 2)	0.09	0.15	0.13	0.08	0.34	0.12	0.29
MixNormal(3, 3)	0.08	0.10	0.10	0.10	0.33	0.16	0.31
MixNormal(4, 4)	0.05	0.12	0.08	0.08	0.29	0.17	0.33
MixGED(2, 2)	0.07	0.18	0.11	0.08	0.30	0.22	0.20
MixGED(3, 2)	0.08	0.15	0.12	0.09	0.34	0.10	0.34
MixGED(3, 3)	0.07	0.11	0.14	0.13	0.32	0.20	0.33
MixGED(4, 4)	0.07	0.12	**0.08**	0.11	0.29	0.13	0.36
A^1MixStable(2, 2)	**0.04**	0.17	0.20	0.12	0.33	0.10	0.21
A^1MixStable(3, 2)	0.05	0.16	0.14	0.09	0.34	0.13	0.26
A^1MixStable(3, 3)	0.06	0.16	0.14	0.10	0.30	0.15	0.25
A^1MixStable(4, 4)	0.07	**0.07**	0.10	0.07	0.28	0.21	0.32
A^2MixStable(2, 2)	0.18	0.09	0.09	0.07	0.11	0.07	0.26
A^2MixStable(3, 2)	0.19	0.08	0.10	0.08	**0.10**	**0.05**	0.27
A^2MixStable(3, 3)	0.19	0.11	0.10	**0.07**	0.11	0.05	0.28
A^2MixStable(4, 4)	0.19	0.08	0.09	0.07	0.11	0.05	0.27
5% VaR							
MixNormal(2, 2)	0.19	0.36	0.22	0.31	**0.34**	0.53	**0.20**
MixNormal(3, 2)	0.17	0.25	0.16	0.28	0.47	0.69	0.49
MixNormal(3, 3)	**0.10**	0.23	0.10	0.28	0.47	0.62	0.53
MixNormal(4, 4)	0.20	0.22	0.10	0.27	0.42	0.59	0.64
MixGED(2, 2)	0.20	0.27	0.16	0.31	0.40	**0.31**	0.35
MixGED(3, 2)	0.19	0.29	0.12	0.26	0.46	0.54	0.60
MixGED(3, 3)	0.16	0.23	0.12	0.28	0.47	0.33	0.61
MixGED(4, 4)	0.17	0.22	0.10	0.27	0.44	0.37	0.64
A^1MixStable(2, 2)	0.25	0.26	0.14	0.31	0.42	0.54	0.28
A^1MixStable(3, 2)	0.16	0.24	0.13	0.29	0.45	0.66	0.40
A^1MixStable(3, 3)	0.16	0.18	0.10	0.30	0.49	0.60	0.56
A^1MixStable(4, 4)	0.17	**0.16**	**0.09**	0.28	0.42	0.63	0.65
A^2MixStable(2, 2)	0.30	0.22	0.17	0.25	0.55	0.34	0.31
A^2MixStable(3, 2)	0.30	0.19	0.18	0.26	0.58	0.41	0.35
A^2MixStable(3, 3)	0.31	0.18	0.21	**0.24**	0.57	0.42	0.33
A^2MixStable(4, 4)	0.31	0.19	0.20	0.24	0.59	0.41	0.31
10% VaR							
MixNormal(2, 2)	0.28	0.53	0.45	0.49	**0.30**	0.65	0.59
MixNormal(3, 2)	0.25	0.43	0.44	0.43	0.42	0.78	0.36
MixNormal(3, 3)	**0.17**	0.35	0.40	0.41	0.48	0.83	0.39
MixNormal(4, 4)	0.23	0.29	0.36	0.39	0.45	0.82	0.52
MixGED(2, 2)	0.28	0.44	0.41	0.52	0.36	0.53	0.27
MixGED(3, 2)	0.26	0.50	0.38	0.46	0.42	0.70	0.49
MixGED(3, 3)	0.22	0.36	0.42	0.43	0.48	0.55	0.50
MixGED(4, 4)	0.21	0.37	0.40	0.40	0.45	0.60	0.52
A^1MixStable(2, 2)	0.31	0.33	**0.27**	0.51	0.38	0.66	**0.25**
A^1MixStable(3, 2)	0.21	0.36	0.35	**0.38**	0.41	0.68	0.30
A^1MixStable(3, 3)	0.20	0.28	0.32	0.43	0.46	0.84	0.44
A^1MixStable(4, 4)	0.21	**0.23**	0.30	0.42	0.44	0.86	0.52
A^2MixStable(2, 2)	0.69	0.49	0.46	0.52	0.46	**0.47**	0.73
A^2MixStable(3, 2)	0.68	0.48	0.46	0.51	0.47	0.48	0.71
A^2MixStable(3, 3)	0.72	0.43	0.45	0.51	0.48	0.47	0.70
A^2MixStable(4, 4)	0.73	0.47	0.45	0.52	0.49	0.48	0.76

Table 1.9: Integrated root mean squared error of the VaR prediction up to the 1%, 5% and 10% level for all models under study (with $\delta = 1$). Entries in boldface denote the best estimate. Results are based on the same out-of-sample exercise as in Table 1.3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Unconditional Coverage, LR_{UC}							
MixNormal(2, 2)	**3.1e-4**	0.14	0.88	0.68	3.23*	0.28	**0.05**
MixNormal(3, 2)	0.05	0.14	0.03	0.68	4.14**	**0.07**	5.62**
MixNormal(3, 3)	1.07	0.14	0.32	1.07	4.14**	0.28	6.81***
MixNormal(4, 4)	0.39	0.32	0.03	0.39	4.14**	0.28	5.62**
MixGED(2, 2)	**3.1e-4**	0.14	0.32	0.39	5.16**	1.70	1.07
MixGED(3, 2)	0.05	0.56	0.14	0.68	4.14**	**0.07**	4.57**
MixGED(3, 3)	0.68	0.14	**3.1e-4**	1.56	5.16**	1.11	5.62**
MixGED(4, 4)	0.68	0.14	**3.1e-4**	0.68	5.16**	0.28	4.57**
A^1MixStable(2, 2)	0.32	0.32	0.14	0.68	5.16**	0.28	3.65*
A^1MixStable(3, 2)	0.14	0.56	**3.1e-4**	0.68	3.23*	**0.07**	4.57**
A^1MixStable(3, 3)	**3.1e-4**	0.32	0.56	0.68	4.14**	0.28	2.84*
A^1MixStable(4, 4)	0.68	**0.03**	**3.1e-4**	0.39	4.14**	0.28	3.65*
A^2MixStable(2, 2)	1.56	0.05	0.03	0.39	**1.70**	**0.07**	2.84*
A^2MixStable(3, 2)	1.56	0.05	0.14	0.39	**1.70**	**0.07**	3.65*
A^2MixStable(3, 3)	1.56	0.05	0.03	**0.05**	**1.70**	**0.07**	4.57**
A^2MixStable(4, 4)	1.56	0.05	0.03	0.17	**1.70**	**0.07**	3.65*
Independence, LR_{IND}							
MixNormal(2, 2)	1.28	0.63	0.77	0.39	1.06	0.37	0.50
MixNormal(3, 2)	1.40	0.63	0.59	0.39	0.92	**0.33**	0.19
MixNormal(3, 3)	1.98	0.63	0.67	0.36	0.92	0.37	**0.16**
MixNormal(4, 4)	1.67	0.67	0.59	0.43	0.92	0.37	0.19
MixGED(2, 2)	1.28	0.63	0.67	0.43	0.79	0.53	0.36
MixGED(3, 2)	1.40	0.72	0.63	0.39	0.92	**0.33**	0.21
MixGED(3, 3)	1.82	0.63	**0.54**	**0.32**	0.79	0.48	0.19
MixGED(4, 4)	1.82	0.63	**0.54**	0.39	0.79	0.37	0.21
A^1MixStable(2, 2)	**0.96**	0.67	0.63	0.39	0.79	0.37	0.24
A^1MixStable(3, 2)	1.06	0.72	**0.54**	0.39	1.06	**0.33**	0.21
A^1MixStable(3, 3)	1.28	0.67	0.72	0.39	0.92	0.37	0.26
A^1MixStable(4, 4)	1.82	0.59	**0.54**	0.43	0.92	0.37	0.24
A^2MixStable(2, 2)	2.15	**0.50**	0.59	0.43	1.37	**0.33**	0.26
A^2MixStable(3, 2)	2.15	**0.50**	0.63	0.43	**0.53**	**0.33**	0.24
A^2MixStable(3, 3)	2.15	**0.50**	0.59	0.50	1.37	**0.33**	0.21
A^2MixStable(4, 4)	2.15	**0.50**	0.59	0.46	1.37	**0.33**	0.24
Conditional Coverage, LR_{CC}							
MixNormal(2, 2)	1.28	0.77	1.65	1.08	4.28	0.66	**0.55**
MixNormal(3, 2)	1.45	0.77	0.62	1.08	5.06	**0.40**	5.81
MixNormal(3, 3)	3.05	0.77	0.99	1.43	5.06	0.66	6.97*
MixNormal(4, 4)	2.06	0.99	0.62	0.81	5.06	0.66	5.81
MixGED(2, 2)	1.28	0.77	0.99	0.81	5.95	2.24	1.43
MixGED(3, 2)	1.45	1.29	0.77	1.08	5.06	**0.40**	4.78
MixGED(3, 3)	2.50	0.77	**0.54**	1.89	5.95	1.58	5.81
MixGED(4, 4)	2.50	0.77	**0.54**	1.08	5.95	0.66	4.78
A^1MixStable(2, 2)	1.28	0.99	0.77	1.08	5.95	0.66	3.88
A^1MixStable(3, 2)	**1.20**	1.29	**0.54**	1.08	4.28	**0.40**	4.78
A^1MixStable(3, 3)	1.28	0.99	1.29	1.08	5.06	0.66	3.11
A^1MixStable(4, 4)	2.50	0.62	**0.54**	0.81	5.06	0.66	3.88
A^2MixStable(2, 2)	3.71	**0.55**	0.62	0.81	3.08	**0.40**	3.11
A^2MixStable(3, 2)	3.71	**0.55**	0.77	0.81	**2.24**	**0.40**	3.88
A^2MixStable(3, 3)	3.71	**0.55**	0.62	**0.55**	3.08	**0.40**	4.78
A^2MixStable(4, 4)	3.71	**0.55**	0.62	0.64	3.08	**0.40**	3.88

Table 1.10: Test statistics at the 1%-VaR level, $LR_{CC} = LR_{UC} + LR_{IND}$, as described in Christoffersen (1998) for all models under study (with $\delta = 1$). Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 1.3.

Bibliography

Adler, R. J. (1997). Discussion: Heavy Tail Modeling and Teletraffic Data. *The Annals of Statistics*, 25(5):1849–1852.

Alexander, C. and Lazar, E. (2006). Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling. *Journal of Applied Econometrics*, 21:307–336.

Alexander, C. and Lazar, E. (2009). Modelling Regime-Specific Stock Price Volatility. *Oxford Bulletin of Economics and Statistics*, 71(6):761–797.

Ausín, M. C. and Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, 51(5):2636–2652.

Badescu, A., Kulperger, R., and Lazar, E. (2008). Option Valuation with Normal Mixture GARCH Models. *Studies in Nonlinear Dynamics & Econometrics*, 12(2):5.

Bai, X., Russell, J. R., and Tiao, G. C. (2003). Kurtosis of GARCH and Stochastic Volatility Models with Non–Normal Innovations. *Journal of Econometrics*, 114:349–360.

Bauwens, L., Hafner, C. M., and Rombouts, J. V. K. (2007). Multivariate Mixed Normal Conditional Heteroskedasticity. *Computational Statistics & Data Analysis*, 51(7):3551–3566.

Bauwens, L., Preminger, A., and Rombouts, J. (2010). Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, 13(2):218–244.

Bauwens, L. and Storti, G. (2009). A Component GARCH Model with Time Varying Weights. *Studies in Nonlinear Dynamics & Econometrics*, 13(2):Article 1.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bonato, M. (2011). Modeling fat tails in stock returns: a multivariate stable-GARCH approach. *Computational Statistics*, 26:1–23.

Broda, S. A. (2011). Tail Probabilities and Partial Moments for Quadratic Forms in Multivariate Generalized Hyperbolic Random Vectors. Working Paper, University of Amsterdam.

Broda, S. A. and Paolella, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of Financial Econometrics*, 7(4):412–436.

Broda, S. A. and Paolella, M. S. (2011). Expected Shortfall for Distributions in Finance. In Čížek, P., Härdle, W., and Rafał Weron, editors, *Statistical Tools for Finance and Insurance*.

Brooks, R. D., Faff, R. W., McKenzie, M. D., and Mitchell, H. (2000). A Multi-country Study of Power ARCH Models and National Stock Market Returns. *Journal of International Money and Finance*, 19(3):377–397.

Chen, Y., Härdle, W., and Spokoiny, V. (2006). GHICA — Risk Analysis with GH Distributions and Independent Components. SFB 649 Discussion Paper 2006-78, Humboldt-University, Berlin.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Day, N. E. (1969). Estimating the Components of a Mixture of Normal Distributions. *Biometrika*, 56(3):463–474.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106.

Doganoglu, T., Hartz, C., and Mittnik, S. (2007). Portfolio Optimization when Risk Factors are Conditionally Varying and Heavy Tailed. *Computational Economics*, 29(3-4):333–354.

Doganoglu, T. and Mittnik, S. (1998). An Approximation Procedure for Asymmetric Stable Paretian Densities. *Computational Statistics*, 13:463–475.

Dowd, K. (2005). *Measuring Market Risk*. Wiley, New York, 2nd edition.

Fabricius, T., Kidmose, P., and Hansen, L. (2001). Dynamic components of linear stable mixtures from fractional low order moments. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 6:3957–3960.

Fama, E. F. (1965). The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105.

Fan, J., Wang, M., and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:679–702.

Francq, C., Roussignol, M., and Zakoïan, J.-M. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22(2):197–220.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.

Giacometti, R., Bertocchi, M. I., Rachev, S. T., and Fabozzi, F. J. (2007). Stable Distributions in the Black-Litterman Approach to Asset Allocation. *Quantitative Finance*, 7(4):423–433.

Gil-Pelaez, J. (1951). Note on the Inversion Theorem. *Biometrika*, 38:481–482.

Giot, P. and Laurent, S. (2003). Value-at-Risk for Long and Short Trading Positions. *Journal of Applied Econometrics*, 18(6):641–663.

Haas, M., Mittnik, S., and Paolella, M. S. (2004a). A New Approach to Markov Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Haas, M., Mittnik, S., and Paolella, M. S. (2004b). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.

Haas, M., Mittnik, S., and Paolella, M. S. (2009). Asymmetric Multivariate Normal Mixture GARCH. *Computational Statistics & Data Analysis*, 53(6):2129–2154.

Haas, M. and Paolella, M. S. (2011). Mixture and Regime-switching GARCH Models. In Bauwens, L., Hafner, C., and Laurent, S., editors, *Handbook of Volatility Models and Their Applications*. Wiley.

Hathaway, R. J. (1985). A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions. *The Annals of Statistics*, 13(2):795–800.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Annals of Statistics*, 3(5):1163–1174.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley-Interscience, 1 edition.

Jansen, D. W. and de Vries, C. G. (1991). On the Frequency of Large Stock Returns: Putting Booms and Busts Into Perspective. *Review of Economics and Statistics*, 73:18–24.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62(1):49–66.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *The Annals of Mathematical Statistics*, 27(4):887–906.

Kiefer, N. M. (1978). Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model. *Econometrica*, 46(2):427–434.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J. (2008). Financial Market Models with Lévy Processes and Time-Varying Volatility. *Journal of Banking & Finance*, 32(7):1363–1378.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J. (2009a). Computing VaR and AVaR In Infinitely Divisible Distributions. Yale ICF Working Paper 09-07, Yale School of Management.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J. (2010). Tempered Stable and Tempered Infinitely Divisible GARCH Models. *Journal of Banking & Finance*, 34(9):2096–2109.

Kim, Y. S., Rachev, S. T., Chung, D. M., and Bianchi, M. L. (2009b). The Modified Tempered Stable Distribution, GARCH Models, and Option Pricing. *Probability and Mathematical Statistics*, 29:91–117.

Kirchler, M. and Huber, J. (2007). Fat Tails and Volatility Clustering in Experimental Asset Markets. *Journal of Economic Dynamics & Control*, 31:1844–1874.

Knight, J. and Satchell, S. (2001). *Return Distributions in Finance*. Butterworth-Heinemann, Oxford.

Kon, S. J. (1984). Models of Stock Returns: A Comparison. *The Journal of Finance*, 39(1):147–165.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press, Cambridge.

Kratz, M. and Resnick, S. I. (1996). The QQ-Estimator and Heavy Tails. *Communications in Statistics – Stochastic Models*, 12(4):699–724.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Lejeune, B. (2009). A Diagnostic m-test for Distributional Specification of Parametric Conditional Heteroscedasticity Models for Financial Data. *Journal of Empirical Finance*, 16(3):507–523.

Liu, J.-C. (2007). Stationarity for a Markov–switching Box–Cox Transformed Threshold GARCH Process. *Statistics and Probability Letters*, 77:1428–1438.

Lombardi, M. J. and Veredas, D. (2009). Indirect estimation of elliptical stable distributions. *Computational Statistics & Data Analysis*, 53(6):2309–2324.

Lyness, J. N. (1969). Notes on the adaptive Simpson quadrature routine. *Journal of the Association for Computing Machinery*, 16(3):483–495.

Malevergne, Y., Pisarenko, V., and Sornette, D. (2005). Empirical Distributions of Stock Returns: Between the Stretched Exponential and the Power Law? *Quantitative Finance*, 5:379–401.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business*, 36:394–419.

McCulloch, J. H. (1985). Interest-Risk Sensitive Deposit Insurance Premia: Stable ACH Estimates. *Journal of Banking and Finance*, 9:137–156.

McCulloch, J. H. (1996). Financial Applications of Stable Distributions. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics*, volume 14. Elsevier Science.

McCulloch, J. H. (1997). Measuring Tail Thickness in Order to Estimate the Stable Index α: A Critique. *Journal of Business and Economic Statistics*, 15:74–81.

McCulloch, J. H. (1998). Numerical Approximation of the Symmetric Stable Distribution and Density. In Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors, *A Practical Guide to Heavy Tails*, pages 489–499. Birkhäuser, Boston.

Mercuri, L. (2008). Option Pricing in a GARCH Model with Tempered Stable Innovations. *Finance Research Letters*, 5(3):172–182.

Mittnik, S., Doganoglu, T., and Chenyao, D. (1999). Computing the Probability Density Function of the Stable Paretian Distribution. *Mathematical and Computer Modelling*, 29:235–240.

Mittnik, S. and Paolella, M. S. (2003). Prediction of Financial Downside Risk with Heavy Tailed Conditional Distributions. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*. Elsevier Science, Amsterdam.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (1998). A Tail Estimator for the Index of the Stable Paretian Distribution. *Communications in Statistics–Theory and Methods*, 27(5):1239–1262.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (2002). Stationarity of Stable Power–GARCH Processes. *Journal of Econometrics*, 106:97–107.

Nolan, J. P. (1997). Numerical Calculation of Stable Densities and Distribution Functions. *Stochastic Models*, 13(4):759–774.

Nolan, J. P. (1998). Univariate Stable Distributions: Parameterizations and Software. In Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors, *A Practical Guide to Heavy Tails*, pages 527–533. Birkhäuser, Boston.

Nolan, J. P. (2012). *Stable Distributions – Models for Heavy Tailed Data*. Birkhäuser, Boston. forthcoming; Chapter 1 online.

Paolella, M. S. (2001). Testing the Stable Paretian Assumption. *Mathematical and Computer Modelling*, 34:1095–1112.

Paolella, M. S. (2007). *Intermediate Probability: A Computational Approach*. John Wiley & Sons, Chichester.

Paolella, M. S. and Steude, S. C. (2008). Risk Prediction: A DWARF-like Approach. *Journal of Risk Model Validation*, 2(1):25–43.

Platen, E. and Heath, D. (2006). *A Benchmark Approach to Quantitative Finance*. Springer, Berlin, Germany.

Poirot, J. and Tankov, P. (2006). Monte Carlo Option Pricing for Tempered Stable (CGMY) Processes. *Asia-Pacific Financial Markets*, 13(4):327–344.

Rachev, S. T. (2003). *Handbook of Heavy Tailed Distributions in Finance*. North-Holland.

Rachev, S. T. and Mittnik, S. (2000). *Stable Paretian Models in Finance*. Wiley & Sons, Chichester.

Rombouts, J. V. K. and Bouaddi, M. (2009). Mixed Exponential Power Asymmetric Conditional Heteroskedasticity. *Studies in Nonlinear Dynamics & Econometrics*, 13(3):Article 3.

Rombouts, J. V. K. and Stentoft, L. (2009). Bayesian Option Pricing Using Mixed Normal Heteroskedasticity Models. CREATES Research Papers 2009-07, School of Economics and Management, University of Aarhus.

Salas-Gonzalez, D., Kuruoglu, E. E., and Ruiz, D. P. (2009). Finite Mixtures of α-Stable Distributions. *Digital Signal Processing*, 19(2):250–264.

Samanidou, E., Zschischang, E., and Stauffer, D. (2007). Agent-Based Models of Financial Markets. *Reports on Progress in Physics*, 70:409–450.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, London.

Seneta, E. (2004). Fitting the Variance-Gamma Model to Financial Data. *Journal of Applied Probability*, 41, Stochastic Methods and Their Applications:177–187.

Stoyanov, S., Samorodnitsky, G., Rachev, S., and Ortobelli, S. (2006). Computing the Portfolio Conditional Value-at-Risk in the alpha-stable Case. *Probability and Mathematical Statistics*, 26:1–22.

Sy, W. (2006). On the Coherence of VaR Risk Measure for Levy Stable Distributions. Working papers, Australian Prudential Regulation Authority.

Tanaka, K. (2009). Strong Consistency of the Maximum Likelihood Estimator for Finite Mixtures of Location–Scale Distributions When Penalty is Imposed on the Ratios of the Scale Parameters. *Scandinavian Journal of Statistics*, 36:171–184.

Vigfusson, R. (1997). Switching between Chartists and Fundamentalists: A Markov Regime–Switching Approach. *International Journal of Finance and Economics*, 2:291–305.

Vlaar, P. J. G. and Palm, F. C. (1993). The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps. *Journal of Business and Economic Statistics*, 11(3):351–360.

Weron, R. (2001). Levy-stable Distributions Revisited: Tail Index > 2 does not exclude the Levy-stable Regime. *International Journal of Modern Physics C*, 12:209–223.

Zolotarev, V. M. (1986). *One-Dimensional Stable Laws*. American Mathematical Society. Translation of Odnomernye Ustoichivye Raspredeleniia, NAUKA, Moskow, 1983.

Chapter 2

Time-varying Mixture GARCH Models and Asymmetric Volatility

Time-varying Mixture GARCH Models and Asymmetric Volatility*,†

Markus Haas[a] Jochen Krause[b] Marc S. Paolella[b,c] Sven C. Steude[b]

[a]*Institute for Quantitative Business and Economics Research, University of Kiel, Germany*

[b]*Department of Banking and Finance, University of Zurich, Switzerland*

[c]*Swiss Finance Institute*

Abstract

The class of mixed normal conditional heteroskedastic (MixN-GARCH) models, which couples a mixed normal distributional structure with GARCH-type dynamics, has been shown to offer a plausible decomposition of the contributions to volatility, as well as excellent out-of-sample forecasting performance, for financial asset returns. In this paper, we generalize the MixN-GARCH model by relaxing the assumption of constant mixing weights. Two different specifications with time–varying mixing weights are considered. In particular, by relating current weights to past returns and realized (component-wise) likelihood values, an empirically reasonable representation of Engle and Ng's (1993) news impact curve with an asymmetric impact of unexpected return shocks on future volatility is obtained. An empirical out-of-sample study confirms the usefulness of the new approach and gives evidence that the leverage effect in financial returns data is closely connected, in a non-linear fashion, to the time–varying interplay of mixture components representing, for example, various groups of market participants.

Keywords — GARCH; News Impact Curve; Leverage Effect; Down-Market Effect; Mixtures; Time–Varying Weights; Value–at–Risk.

JEL classification: C22; C51; G10

*This paper is a highly altered and extended version of the manuscript Haas et al. (2006b). The authors are grateful to Emese Lazar, Michael McAleer, and an anonymous referee, for discussions and comments which led to a substantially improved paper. The research of Haas was supported by the *Deutsche Forschungsgemeinschaft*. Part of the research of Paolella and Steude has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation.

†A condensed version was published in The North American Journal of Economics and Finance, 2013, Vol. 26, pp. 602–623.

2.1 Introduction

Among the many apparent empirical regularities of financial time series, one of the most researched is the relationship between equity returns and volatility. Over the last two decades, a large amount of literature reports an asymmetric volatility response between positive and negative returns.

The initial discovery of asymmetry in the relationship between returns and volatility is usually accredited to Black (1976) and Christie (1982) with their observation that current returns and future volatility are negatively correlated, commonly referred to in the literature as Black's leverage effect. The historic explanations of such market behavior are grounded in the firms' dept–equity ratio that changes with movements in the return and, thus, alters the stock's riskiness. However, an increasing number of studies challenge this fundamental reasoning. For example, Hasanhodzic and Lo (2011) find the leverage effect also present in all–equity financed companies and report its effect even stronger than for leveraged firms. Similarly, Hens and Steude (2009) find the effect in the laboratory environment absent of any leverage implying that the inverse relationship between price and volatility is not driven by financial leverage. In addition, Figlewski and Wang (2000) present evidence that the leverage effect is largely independent of a change in the firms' capital structure. More evidence for the leverage effect in assets for which the traditional explanation cannot hold is provided in Park (2011), who conjectures a herding type of behavior to explain it.

Insight into the asymmetric volatility response has also risen from a different strand of literature. The ARCH and GARCH model classes – that in their original version are such that negative and positive return shocks have the *same* impact on volatilities – have been extended by several authors to include asymmetric effects as well, e.g., in the univariate case, models that allow for this effect include the EGARCH model of Nelson (1991), the GJR model of Glosten et al. (1993) and the threshold ARCH model of Zakoian (1994). All of these volatility models are asymmetric in a sense that "bad" news tend to be associated with a larger increase in (tomorrow's) volatility than "good" news of the same magnitude – but positive news does not *reduce* volatility like the leverage effect suggests; compare Asai and McAleer (2011). Only the EGARCH model, although not guaranteeing the leverage effect, permits the effect subject to restrictions on the size and sign parameters.

Some elements of the structure in the return-volatility relationship are not fully understood yet and there is still a lively debate about its dynamics. In this paper, we propose a mixture GARCH approach that can represent a variety of different asymmetric response patterns. The model yields a new and flexible dynamic structure for modeling the (generally) asymmetric relationship between returns and volatility that allows feedback between different components of variances and the overall process. The goal of this paper is to study these volatility dynamics in detail. Our proposed model has a rich GARCH structure, so that an increase in volatility occurs when a negative or positive shock hits the market, but its impact is enhanced for negative shocks, while mitigated for positive shocks.

Further, the use of a mixed normal distribution for modeling the unconditional distribution of asset returns has been considered by numerous authors, including Fama (1965), Kon (1984), Tucker and Pond (1988), and Aparicio and Estrada (2001). More recently, Kim and White (2004, p. 72) provide further evidence of the appropriateness of normal mixtures for financial data, stating "[We propose that] it may be more productive to think of the S&P500 index returns studied here as being better described as a mixture containing a predominant component that is nearly symmetric with mild kurtosis and a relatively rare component that generates highly extreme behavior." Along similar lines, Neftci (2000) argues that the extreme movements in asset prices are caused by mechanisms which are "structurally different" from the "routine functioning of markets".

The problem with any unconditional model for asset returns is that they cannot capture the blatant volatility clustering inherent in virtually all return series observed at weekly or higher frequencies, and will suffer appropriately in terms of short–term Value–at–Risk (VaR) forecasting ability. The effectiveness and easy implementation of GARCH models for this purpose is undisputed, and numerous variations and extensions of Bollerslev's (1986) original construct have been proposed and shown to deliver superior forecasts; see, for example, Palm (1996), Kuester et al. (2006), and Alexander (2008, Ch. 4) for surveys.

The mixed normal GARCH, or MixN-GARCH, is a relatively recent GARCH–type model class which combines the features of normal mixture distributions and a GARCH model, and has been independently proposed and investigated by Alexander and Lazar (2006) and Haas et al. (2004a,b). By judiciously coupling a k–mixture of normal distributions with a GARCH–type dynamic structure that links the k density components, several previously advocated models are nested, and a variety of stylized facts of asset returns can be successfully modeled, such as the usual fat tails and volatility clustering, but also time–varying skewness and kurtosis. The model has been shown in the aforementioned papers to offer a plausible decomposition of the contributions to market volatility, and also to deliver highly competitive out-of-sample forecasts. For further detail and more recent extensions, see Haas and Paolella (2012).

A common property of MixN-GARCH is the constancy of mixing weights of the component densities, which often allows for a straightforward interpretation of the impact of the individual components. However, constancy of the distributional proportions is not necessarily a realistic assumption, and, as we demonstrate below, leads to less accurate forecasts compared with a more general class of models which does allow for time–varying mixture weights.

While the use of mixtures, in particular, the mixture of normals distribution, is ubiquitous in numerous scientific applications, of which finance is only one of many examples (see, e.g., McLachlan and Peel, 2000; and Frühwirth-Schnatter, 2010), the thorny, and very real, issue of avoiding the singularities when maximizing the likelihood needs to be addressed. We employ the new, easily implemented, and theoretically very attractive method introduced in Broda et al. (2013), which is applicable to unconditional mixtures, as well as mixture-GARCH models. This renders model estimation to be very simple to implement, as fast as standard likelihood optimization, numerically fully unproblematic, and, under appropriate conditions on the data generation process, the resulting maximum likelihood estimates are consistent.

Anticipating the empirical results in Section 2.4, the newly proposed model, denoted MixN-GARCH-LIK, performs very well according to numerous out-of-sample criteria for the majority of the considered data (seven major equity indices and exchange rates). Given the ease of use in implementation and estimation, as well as the general appeal of mixture distributions in finance, from both economic and empirical perspectives, we show that the new class of models provide a worthy contribution for forecasting the distribution and tail risk of univariate financial returns data.

The remainder of this paper is as follows. Section 2.2 briefly introduces the MixN-GARCH model. Section 2.3 discusses its extension to allow for time–varying mixing weights and reviews the implied news impact curves. Section 2.4 details an empirical exercise, and Section 2.5 concludes.

2.2 Mixed Normal GARCH

In the mixed normal GARCH (MixN-GARCH) model the conditional density of return r_t is assumed to be a *finite normal mixture distribution* with k components. That is, with f_t denoting a conditional density based on the information set at time t,

$$f_{t-1}(r_t; \lambda_{1t}, \ldots, \lambda_{k,t}, \mu_{1t}, \ldots, \mu_{k,t}, \sigma_{1t}^2, \ldots, \sigma_{k,t}^2) = \sum_{j=1}^{k} \lambda_{jt} \phi(r_t; \mu_{jt}, \sigma_{jt}^2), \tag{2.1}$$

where

$$\phi(r_t; \mu_{jt}, \sigma_{jt}^2) = \frac{1}{\sqrt{2\pi}\sigma_{jt}} \exp\left\{ -\frac{(r_t - \mu_{jt})^2}{2\sigma_{jt}^2} \right\}$$

is the normal density, the strictly positive *mixing weights* (or *probabilities*) λ_{jt} satisfy $\sum_j \lambda_{jt} = 1$, and the $k \times 1$ vector $\boldsymbol{\sigma}_t^{(2)} = (\sigma_{1t}^2, \ldots, \sigma_{k,t}^2)'$ of conditional *component variances* follows a GARCH(p, q) process of the form

$$\boldsymbol{\sigma}_t^{(2)} = \boldsymbol{\omega} + \sum_{i=1}^{q} \boldsymbol{\alpha}_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \boldsymbol{\beta}_i \boldsymbol{\sigma}_{t-i}^{(2)}, \tag{2.2}$$

where the error term

$$\varepsilon_t = r_t - \mathbb{E}_{t-1}[r_t] = r_t - \sum_{j=1}^{k} \lambda_{jt}\mu_{jt}, \tag{2.3}$$

and $\boldsymbol{\omega} \in \mathbb{R}^k$, $\boldsymbol{\alpha}_i \in \mathbb{R}^k$, $i = 1, \ldots, q$, and $\boldsymbol{\beta}_i \in \mathbb{R}^{k \times k}$, $i = 1, \ldots, p$, are parameters matrices which have to obey restrictions to guarantee that $\boldsymbol{\sigma}_t^{(2)}$ remains positive for all t. As with the standard (single-component) GARCH model, $p = q = 1$ is typically found to be sufficient; moreover, the *diagonal* GARCH specification with a diagonal $\boldsymbol{\beta}_1$ is typically favored in empirical applications and admits a clear-cut interpretation of the component-specific volatility processes (see Haas et al., 2004b; and Haas and Paolella, 2012, Section 3.2.3, for discussion). In this case, we write the model as

$$\boldsymbol{\sigma}_t^{(2)} = \boldsymbol{\omega} + \boldsymbol{\alpha}\varepsilon_{t-1}^2 + \boldsymbol{\beta}\boldsymbol{\sigma}_{t-1}^{(2)}, \tag{2.4}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)' > \mathbf{0}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)' \geq \mathbf{0}$, and $\boldsymbol{\beta} = \text{diag}(\beta_1, \ldots, \beta_k) \geq \mathbf{0}$, where the inequalities have to hold element-wise.

The conditional mean of r_t has already been introduced in (2.3). Its conditional variance implied by the mixture density (2.1) is of great interest in the discussion that follows and is given by

$$\mathbb{V}_{t-1}(r_t) = \sum_{j=1}^{k} \lambda_{jt}(\sigma_{jt}^2 + \mu_{jt}^2) - \left(\sum_{j=1}^{k} \lambda_{jt}\mu_{jt} \right)^2. \tag{2.5}$$

Alexander and Lazar (2006) and Haas et al. (2004b) consider the case where the mixing weights, λ_{jt}, and the component means, μ_{jt}, are constant over time, but the generalization in Equations (2.1)–(2.3), with these quantities being time–varying, is conceptually straightforward. In this paper, we consider MixN-GARCH specifications with time–varying weights to capture an asymmetric impact of negative and positive and/or small and large shocks on future volatility, as discussed in the introduction.

2.3 Time-varying weights

The idea of modeling economic variables using mixtures with time–varying mixing weights (or regime probabilities) is not new. Most notably perhaps, the Markov–switching model of Hamilton (1989), which has many applications in macroeconomics and finance, can be interpreted in this framework. In addition, in a number of applications, mixture models with mixing weights depending on lagged process values as well as exogenous variables have been employed quite successfully. An example is the modeling of exchange rate behavior in target zones, where a jump component reflects the probability of realignments, and the probability of a jump depends on interest differentials and, possibly, further explanatory variables incorporating market expectations (see, e.g., Vlaar and Palm, 1993; Bekaert and Gray, 1998; Neely, 1999;

Klaster and Knot, 2002 and Haas et al., 2006a). Cheng et al. (2009) provide an application to national stock index returns, and Tashman and Frey (2009) successfully use such models to capture a nonlinear relation between hedge fund returns and various market risk factors. The conditional densities of such mixture models exhibit an enormous flexibility. For example, as illustrated by Haas et al. (2006a) in an application to the EMS crisis of 1992, the predictive density may become bimodal when the probability of a realignment as well as the expected jump size are sufficiently large; see also Wong and Li (2001) for an example of a bimodal predictive density in an hydrological application.

In this paper, we consider two different approaches to specifying time–varying weights in mixture GARCH models. In the first specification, to be discussed in Section 2.3.1, we let the weights depend on the lagged shock in a logistic fashion. By doing so, an asymmetric response of future volatility to negative and positive shocks is introduced, which is a robust feature of many stock return series. In the second variant, presented in Section 2.3.2, we follow a different approach and determine the conditional weights of the mixture components by their respective most recent explanatory power, as measured by their lagged component-specific likelihood contributions. This also induces a certain degree of asymmetry in the volatility response pattern since, for stock returns, there is also a *contemporaneous* negative relation between return and volatility, i.e., the high–volatility component is also that with the smaller expected return. In the outer parts of the distribution, however, volatility effects dominate, and thus both models represent different (potential) aspects of return volatility dynamics.

2.3.1 Time–varying mixture GARCH with logistic mixing weights

A general approach is to relate the weights of the components to past innovations via logistic response functions. In particular, in a two–component model, to which we restrict attention in this paper,[3] the weight of the first component is given by

$$\lambda_t(\boldsymbol{x}_t) = \frac{\exp\{\boldsymbol{\gamma}'\boldsymbol{x}_t\}}{1 + \exp\{\boldsymbol{\gamma}'\boldsymbol{x}_t\}}, \tag{2.6}$$

with $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_{p-1})'$ being a vector of parameters and $\boldsymbol{x}_t$ a vector of p predetermined variables, typically including a constant.

A mixture GARCH model in this vein was considered by Bauwens et al. (2006), who use $\boldsymbol{x}_t = (1, \varepsilon_{t-1}^2)'$, i.e.,

$$\lambda_t(\varepsilon_{t-1}) = \frac{\exp\{\gamma_0 + \gamma_1\varepsilon_{t-1}^2\}}{1 + \exp\{\gamma_0 + \gamma_1\varepsilon_{t-1}^2\}}, \tag{2.7}$$

[3] For possible generalizations to k components, see Haas et al. (2006b).

so that, if $\gamma_1 > 0$, $\lambda_t \to 1$ as ε_{t-1}^2 becomes large. The motivation for this specification is that "large shocks have the effect of relieving pressure by reducing the probability of a large shock in the next period".[4]

Note that, in (2.7), λ_t is a symmetric function of ε_{t-1}, and $\lambda_t(-\infty) = \lambda_t(\infty) = 1$. In this paper, we aim at modeling an asymmetric effect of past shocks, and thus we let $\boldsymbol{x}_t = (1, \varepsilon_{t-1})'$ in (2.6), i.e., we specify the conditional mixing weight as

$$\lambda_t(\varepsilon_{t-1}) = \frac{\exp\{\gamma_0 + \gamma_1 \varepsilon_{t-1}\}}{1 + \exp\{\gamma_0 + \gamma_1 \varepsilon_{t-1}\}}, \tag{2.8}$$

which, when coupled with the MixN-GARCH structure described in Section 2.2, will be referred to as MixN-GARCH-LOG model. The logic behind (2.8) is as follows. Suppose that the first component is the high–volatility regime, and $\gamma_1 > 0$. Then $\mathrm{d}\lambda_\mathrm{t}/\mathrm{d}\varepsilon_{\mathrm{t}-1} = \gamma_1 \lambda_\mathrm{t}(1 - \lambda_\mathrm{t}) > 0$, and, in view of (2.5), the conditional variance will be lower for positive shocks than for negative shocks of the same magnitude. In the next section, the capability of the MixN-GARCH-LOG model to reproduce various asymmetric response patterns is further elucidated via the concept of the news impact curve, as introduced in Engle and Ng (1993).

2.3.1.1 Special cases and relation to other models

In this section, we illustrate the flexibility of the MixN-GARCH-LOG model introduced in Sections 2.2 and 2.3.1 in capturing various asymmetric response patterns of the conditional volatility to previous shocks. We consider various simple special cases of the general model and relate those to some standard asymmetric GARCH models discussed in the literature.[5]

A convenient tool to characterize the impact of news on conditional volatility is the news impact curve (NIC) devised by Engle and Ng (1993). The NIC describes the relation between the conditional variance σ_t^2 and the lagged shock ε_{t-1}, with the lagged conditional variances in the GARCH recursion fixed at their unconditional values. As discussed by Engle and Ng (1993), the NIC of the GARCH is a quadratic function centered at $\varepsilon_{t-1} = 0$, whereas most asymmetric GARCH models have NICs which either still have their minimum at zero but with different slopes for positive and negative shocks or which admit asymmetries by centering the quadratic at a nonzero (usually positive) value. By imposing certain parameter restrictions and considering limiting cases of the MixN-GARCH-LOG model process, we can isolate these typical (and further) shapes of the NIC and thereby get a glimpse of the flexibility of the general (unrestricted) model. To discuss these restrictions, we reproduce the general (non-diagonal)

[4]This assumes that the second component represents the high–volatility regime. In an application to the NASDAQ index, the authors find that, when using (2.7), the evidence for a time–varying mixing weight is weak.

[5]A recent investigation of various asymmetric GARCH specifications is Rodriguez and Ruiz (2012).

MixN-GARCH(1,1)-LOG process for two components, i.e.,

$$\begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \varepsilon_{t-1}^2 + \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{pmatrix}, \tag{2.9}$$

and the weight of the first component is described by (2.8). To fully concentrate on the conditional volatility dynamics, we furthermore assume zero component means, i.e., $\mu_1 = \cdots = \mu_k = 0$.

Consider the situation where, in (2.9), $\beta_{1,2} = \beta_{2,1} = 0$ (diagonal model) and $\alpha_1 = \alpha_2 \equiv \alpha$ and $\beta_{1,1} = \beta_{2,2} \equiv \beta$, i.e., the intercepts differ, whereas the GARCH dynamics are the same in both components. To figure out the NIC for this specification, we observe that it is identical to the one suggested by Vlaar and Palm (1993), where $\sigma_{2t}^2 = \sigma_{1t}^2 + a$ for constant a. To see this, let L denote the lag operator and write the ARCH(∞) representation of σ_{2t}^2 as

$$\sigma_{2t}^2 = \frac{\omega_2}{1-\beta} + \frac{\alpha\varepsilon_{t-1}^2}{1-\beta L} = \frac{\omega_2 - \omega_1}{1-\beta} + \frac{\omega_1}{1-\beta} + \frac{\alpha\varepsilon_{t-1}^2}{1-\beta L} = a + \sigma_{1t}^2, \tag{2.10}$$

where $a = (\omega_2 - \omega_1)/(1-\beta)$. Therefore, from (2.5), the conditional variance is

$$\begin{aligned} \sigma_t^2 &= \lambda_t \sigma_{1t}^2 + (1-\lambda_t)\sigma_{2t}^2 = \lambda_t \sigma_{1t}^2 + (1-\lambda_t)(a + \sigma_{1t}^2) \\ &= (1-\lambda_t)a + \sigma_{1t}^2 = (1-\lambda_t)a + \omega_1 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{1,t-1}^2. \end{aligned} \tag{2.11}$$

If $\gamma_1 > 0$ and $\omega_2 > \omega_1$, λ_t is increasing in ε_{t-1} and $a > 0$, and thus the NIC is asymmetric in that it is higher for negative than for positive shocks of the same magnitude. As in the AGARCH model of Engle (1990) its minimum is also located at a positive value, as can be seen by minimizing (2.11) with respect to ε_{t-1}, i.e., setting to zero

$$\frac{\partial\sigma_t^2}{\partial\varepsilon_{t-1}} = -\frac{\partial\lambda_t}{\partial\varepsilon_{t-1}}a + 2\alpha\varepsilon_{t-1} = -\gamma_1\lambda_t(1-\lambda_t)a + 2\alpha\varepsilon_{t-1},$$

which, as long as $\lambda_t \in (0,1)$, and since $a > 0$ and $\gamma_1 > 0$, can only be zero for a positive ε_{t-1}. The unconditional expectation of σ_{1t}^2 in (2.11) is not known in closed form, but the NIC can still be drawn by evaluating it via simulation and an example is shown in Panel (a) of Figure 2.1.

If we further put $\gamma_0 = 0$ and consider the limiting case $\gamma_1 \to \infty$ in (2.8), so that

$$\lambda_t = \begin{cases} 1, & \text{if } \varepsilon_{t-1} > 0, \\ \frac{1}{2}, & \text{if } \varepsilon_{t-1} = 0, \\ 0, & \text{if } \varepsilon_{t-1} < 0, \end{cases} \tag{2.12}$$

we obtain the sign-switching GARCH model of Fornari and Mele (1997), except that the right-hand side lagged variance in (2.11) is $\sigma_{1,t-1}^2$ rather than the overall variance σ_{t-1}^2. The NIC of this process has a somewhat unusual form; namely it is a "broken" parabola with the positive and negative arms having different intercepts. To calculate the NIC explicitly for this special case (i.e., obtain an explicit expression for $\mathbb{E}[\sigma_{1t}^2]$), we define an indicator variable $\mathbb{1}_t$ which is unity or zero, depending on whether ε_t is drawn from the first or second component, respectively. We can then write, with $\{\eta_t\}$ denoting an iid sequence of standard normal variables,

$$\begin{aligned} \sigma_{1t}^2 &= \omega_1 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{1,t-1}^2 \\ &= \omega_1 + \alpha\eta_{t-1}^2\{\mathbb{1}_{t-1}\sigma_{1,t-1}^2 + (1-\mathbb{1}_{t-1})\sigma_{2,t-1}^2\} + \beta\sigma_{1,t-1}^2 \\ &= \omega_1 + \alpha\eta_{t-1}^2\{\mathbb{1}_{t-1}\sigma_{1,t-1}^2 + (1-\mathbb{1}_{t-1})(\sigma_{1,t-1}^2 + a)\} + \beta\sigma_{1,t-1}^2 \\ &= \omega_1 + \alpha\eta_{t-1}^2(1-\mathbb{1}_{t-1})a + (\alpha\eta_{t-1}^2 + \beta)\sigma_{1,t-1}^2. \end{aligned} \tag{2.13}$$

Thus, provided $\alpha + \beta < 1$, the process is covariance stationary and we have[6]

$$\mathbb{E}[\sigma_{1t}^2] = \frac{\omega_1 + \alpha\, a/2}{1-\alpha-\beta} = \frac{1}{1-\alpha-\beta}\left[\omega_1 + \frac{\omega_2-\omega_1}{2}\frac{\alpha}{1-\beta}\right].$$

An example of such a NIC is shown in Panel (b) of Figure 2.1.

The sign-switching GARCH model is not very successful empirically (Fornari and Mele, 1997) and a more popular asymmetric process is the one of Glosten et al. (1993), i.e., the GJR-GARCH, where the NIC has its minimum at zero but with different slopes for negative and positive shocks. To reproduce such a shape with model (2.9) and (2.8), we may set $\omega_1 = \omega_2 \equiv \omega$, $\beta_{1,1} = \beta_{2,1} = 0$, and $\beta_{1,2} = \beta_{2,2} \equiv \beta$, i.e., we have the restricted *non-diagonal* MixN-GARCH(1,1) model

$$\begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} \omega \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \varepsilon_{t-1}^2 + \begin{pmatrix} 0 & \beta \\ 0 & \beta \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{pmatrix}. \tag{2.14}$$

We may note that, in contrast to the example above, (2.14) does not admit a diagonal GARCH(1,1) representation, as we observe when we plug the ARCH(∞) representation of $\sigma_{2,t-1}^2 = \omega(1-\beta)^{-1} + \alpha_2(1-\beta L)^{-1}\varepsilon_{t-2}^2$ into the equation for σ_{1t}^2, that is,

$$\begin{aligned} \sigma_{1t}^2 &= \omega + \alpha_1\varepsilon_{t-1}^2 + \beta\sigma_{2,t-1}^2 \\ &= \omega + \alpha_1\varepsilon_{t-1}^2 + \frac{\beta\omega}{1-\beta} + \frac{\beta\alpha_2\varepsilon_{t-2}^2}{1-\beta L}, \end{aligned}$$

[6]The unconditional process variance is

$$\mathbb{E}[\varepsilon_t^2] = \frac{1}{2}\mathbb{E}[\sigma_{1t}^2] + \frac{1}{2}\mathbb{E}[\sigma_{2t}^2] = \mathbb{E}[\sigma_{1t}^2] + \frac{a}{2} = \frac{(\omega_1+\omega_2)/2}{1-\alpha-\beta}.$$

which through multiplication with $1-\beta L$ shows that the *diagonal* representation of this process is restricted MixN-GARCH(1,2), namely

$$\begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} \omega \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \varepsilon_{t-1}^2 + \begin{pmatrix} \beta(\alpha_2-\alpha_1) \\ 0 \end{pmatrix} \varepsilon_{t-2}^2 + \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{pmatrix}.$$

Using (2.5), the conditional variance for model (2.14) works out as

$$\sigma_t^2 = \lambda_t \sigma_{1t}^2 + (1-\lambda_t)\sigma_{2t}^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + (1-\lambda_t)(\alpha_2-\alpha_1)\varepsilon_{t-1}^2 + \beta\sigma_{2,t-1}^2. \tag{2.15}$$

With $\alpha_2 > \alpha_1$ and $\gamma_1 > 0$ as above, the NIC is centered at zero but has a larger slope for negative shocks than for positive shocks of the same magnitude, as illustrated in Panel (c) of Figure 2.1.

The limiting case $\gamma_1 \to \infty$ such that (2.12) holds would then correspond to the GJR model, except again that the lagged variance in (2.15) is $\sigma_{2,t-1}^2$ rather than σ_{t-1}^2.[7] Again, for the limiting case, the unconditional expectation of σ_{1t}^2 and hence the NIC can be evaluated exactly. To do so, we define the indicator variable $\mathbb{1}_t$ and $\{\eta_t\}$ as in (2.13), so that we can write

$$\begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} \omega \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_1\eta_{t-1}^2\mathbb{1}_{t-1} & \alpha_1\eta_{t-1}^2(1-\mathbb{1}_{t-1})+\beta \\ \alpha_2\eta_{t-1}^2\mathbb{1}_{t-1} & \alpha_2\eta_{t-1}^2(1-\mathbb{1}_{t-1})+\beta \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{pmatrix}.$$

Thus, the process is covariance stationary if the maximal eigenvalue of the matrix

$$\boldsymbol{C} = \begin{pmatrix} \alpha_1/2 & \alpha_1/2+\beta \\ \alpha_2/2 & \alpha_2/2+\beta \end{pmatrix}$$

is below unity. It follows from the results of Haas et al. (2004b) that this is equivalent to the condition $\beta < 1$ and

$$\det(\boldsymbol{I}_2 - \boldsymbol{C}) = 1 - \beta - \frac{\alpha_1+\alpha_2}{2} - \beta\frac{\alpha_2-\alpha_1}{2} = 1 - \beta - \overline{\alpha} - \beta\frac{\Delta}{2} > 0, \tag{2.16}$$

where $\overline{\alpha} = (\alpha_1+\alpha_2)/2$, and $\Delta = \alpha_2 - \alpha_1$.[8] This condition is not identical (although very similar) to that for the corresponding GJR model, i.e., $\beta + (\alpha_1+\alpha_2)/2 < 1$ (cf. Ling and McAleer, 2002). If condition

[7] We decided to have $\sigma_{2,t-1}^2$ rather than $\sigma_{1,t-1}^2$ appear in the component-specific GARCH recursions in (2.14) due to our assumption that $\alpha_2 > \alpha_1$. It may happen that $\alpha_1 = 0$, i.e., positive shocks have no impact on the conditional volatility. Then, with the roles of $\sigma_{1,t-1}^2$ and $\sigma_{2,t-1}^2$ interchanged in (2.14), i.e., $\beta_{1,1} = \beta_{2,1} \equiv \beta$ and $\beta_{1,2} = \beta_{2,2} = 0$, σ_{1t}^2 would rapidly converge to a constant and σ_{2t}^2 would reduce to an ARCH(1) process.

[8] Note that (2.16) can be rewritten as $\alpha_2 + \beta < 1 + (1-\beta)\Delta/2$, which shows that the GARCH parameters in the second component need not satisfy the condition $\alpha_2 + \beta < 1$. Conditions for stationarity for the general model are not known. The results of Bauwens et al. (2006) cannot be applied since they assume $\lambda_t \to 1$ as $\varepsilon_{t-1}^2 \to \infty$. In such situations, simulation methods as proposed in Gallant et al. (1993) could, in principle, be used to investigate the stationarity of a given model. As far as the properties of the maximum likelihood estimator are concerned, simulations in Cheng et al. (2009) for mixture GARCH models with time-varying weights and typical sample sizes in finance suggest consistency with asymptotic variances being well approximated by the diagonal elements of the inverse of the observed information matrix.

(2.16) holds, the unconditional expectation of the component-specific variances is

$$\mathbb{E}\begin{bmatrix}\sigma_{1t}^2\\ \sigma_{2t}^2\end{bmatrix} = (\boldsymbol{I}_2 - \boldsymbol{C})^{-1}\begin{pmatrix}\omega\\ \omega\end{pmatrix} = \frac{1}{1-\beta-\overline{\alpha}-\beta\Delta/2}\begin{pmatrix}\omega(1-\Delta/2)\\ \omega(1+\Delta/2)\end{pmatrix},$$

which can be used to construct the NIC.

In the unrestricted diagonal MixN-GARCH model, where $\beta_{1,2} = \beta_{2,1} = 0$ and $\beta_{1,1} \equiv \beta_1$ and $\beta_{2,2} \equiv \beta_2$ in (2.14), the conditional variance becomes

$$\begin{aligned}\sigma_t^2 &= \sigma_{1t}^2 + (1-\lambda_t)(\sigma_{2t}^2 - \sigma_{1t}^2) &(2.17)\\ &= \omega_1 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{1,t-1}^2 + \frac{\omega_2 - \omega_1 + (\alpha_2-\alpha_1)\varepsilon_{t-1}^2 + \beta_2\sigma_{2,t-1}^2 - \beta_1\sigma_{1,t-1}^2}{1+\exp\{\gamma_0+\gamma_1\varepsilon_{t-1}\}}. &(2.18)\end{aligned}$$

The simplest possible specification of the form (2.17) appears when *both* conditional regime-specific variances are constant, i.e., in (2.19), $\alpha = \beta = 0$, so that $\sigma_{jt}^2 = \omega_j$, $j = 1, 2$.[9] With $\omega_1 < \omega_2$, this leads to a NIC which decreases monotonically in a logistic fashion, as illustrated in Panel (d) of Figure 2.1. The logistic shape of the NIC is not very reasonable, and a more plausible specification with (potentially) monotonically decreasing NIC is obtained when only one of the component variances is constant, which is termed *partial* MixN-GARCH in Haas et al. (2004b), i.e.,

$$\sigma_{1t}^2 = \omega_1, \quad \sigma_{2t}^2 = \omega_2 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{2,t-1}^2. \tag{2.19}$$

The conditional variance then becomes

$$\sigma_t^2 = \omega_1 + (1-\lambda_t)(\omega_2 - \omega_1 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{2,t-1}^2) = \omega_1 + \frac{\omega_2 - \omega_1 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{2,t-1}^2}{1+\exp\{\gamma_0+\gamma_1\varepsilon_{t-1}\}}. \tag{2.20}$$

As ε_{t-1} increases, (2.20) will eventually converge to ω_1, but the convergence may or may not be monotonic. An example for monotonic convergence (i.e., a monotonically decreasing NIC) is provided in Panel (e) of Figure 2.1. The *unrestricted* diagonal specification with conditional variance as in (2.17) can capture more complex behaviors of conditional volatility. As an example, consider a pattern emphasized by Fornari and Mele (1997), namely that "high negative shocks increase future volatility more than high positive ones while—at the same time—small positive shocks too often produce a stronger impact on future volatility than negative shocks of the same size", as illustrated in Panel (f) of Figure 2.1.[10] This occurs when $\alpha_2 > \alpha_1$ as in the GJR-type model but the NIC assumes its minimum at a negative value.

[9]The model is a standard Gaussian mixture with time–varying weight, a special case of the LMARX process of Wong and Li (2001). As these authors show, such models can conveniently be estimated via a small extension of the EM algorithm for standard iid mixture models.

[10]In Fornari and Mele (1997) the *volatility-switching* GARCH model is designed to reproduce this effect.

2.3.2 Time–varying mixture GARCH with likelihood driven mixing weights

The second model takes on lagged likelihood values as the driver of the current mixing weights. In other words, the time conditional process of the mixing weights is driven by the explanatory power of the (mixture) component models based on their historic performance. We consider this a natural link between yesterday's return and today's volatility as the component model that best explains past returns is rewarded with a higher weight, while the other components proportionally receive lower weights (the vector of mixing weights must sum to one). Different domains of expertise (a term referring to expert systems in the field of cognitive systems in computer science) are thus defined, in a non-linear fashion via the component-wise likelihood functions, by higher and lower mixing weights, which lead to a (possibly) asymmetric news impact curve (NIC) regarding the overall variance of the model. This partitioning is further emphasized by different mixture component means, μ_i. The model structure for the mixing weights is given by

$$\lambda_{jt} = \frac{W_{jt}}{\sum_i W_{it}}, \quad W_{jt} = \nu_j + \sum_{m=1}^{u} \gamma_m \frac{\ell_{j,t-m}}{\sum_i \ell_{i,t-m}}, \tag{2.21}$$

where $\ell_{jt} = \phi\left(r_t; \mu_{jt}, \sigma^2_{jt}\right)$, $\nu_j > 0$, $j = 1, \ldots, k$, and $\gamma_m \geq 0$, $m = 1, \ldots, u$. As in (2.4), a standard GARCH structure is considered for the k mixture components. Similar to (2.6), additional terms (lagged terms of λ_{jt} or of exogenous variables) could be entertained to augment (2.21).[11] We focus, however, on a sparse parametrization and stick to $u = 1$, i.e., $W_{jt} = \nu_j + \gamma \ell_{j,t-1}/(\sum_i \ell_{i,t-1})$. The limiting case for the sparse model, where only $\ell_{p,t-1}$ is different from zero, takes the form

$$\lambda_{jt} = \frac{\nu_j + \gamma \mathbb{1}_{j=p}}{1 + \gamma \mathbb{1}_{j=p}}, \tag{2.22}$$

such that the deviation from $\boldsymbol{\lambda} = \boldsymbol{\nu}$ is bounded above by (2.22) as a function of γ. A leverage-type effect can evolve as a special form of an asymmetric NIC, e.g., compare Asai and McAleer (2011), if the mixture GARCH components (increasingly ordered by their component means) form a decreasing series concerning the amplitudes of their volatility dynamics. To be precise, by construction of the model, this leverage-type effect is limited to the center of the data, where the NIC can be characterized by the different domains of expertise (or regimes of volatility); whereas the GARCH component of highest volatility will dominate the outer area by the scale of γ in (2.22). Hence, one may want to call the modeled effect a *partial leverage effect*, although our empirical testing confirms that the effect (if present) usually affects more than 90% of the observed data. Figure 2.2 shows exemplarily that the effect is indeed found in empirical returns data, while Section 2.4 confirms the usefulness of this approach in an exhaustive out-

[11] Spillover effects, as an integral part of the literature on multivariate (GARCH) models, e.g., see McAleer and da Veiga (2008), may likewise be implemented as in (2.21) by a likelihood driven linkage.

of-sample forecast study. The leverage nature of the effect arises from the change in the mixing weights, such that from a stylized point of view, the high volatility component dominates for (lagged) returns between -3 and 0, while the contrary holds between 0 and 3. Similar patterns are a robust finding in almost all estimates, if $\gamma > 0$.

2.3.2.1 Estimation using an embedded EM Algorithm

We distinguish between two model variants, namely MixN-GARCH-LIK and MixN-GARCH-LIKW. In the first model, MixN-GARCH-LIK, we restrict vector $\boldsymbol{\nu}$ to be estimated separately from the other parameters, and enforce that $\boldsymbol{\nu}$ maximizes the log likelihood function of the MixN-GARCH model without time-varying mixing weight, i.e.,

$$\boldsymbol{\nu} = \arg\max_{\boldsymbol{\lambda}} \sum_{t=1}^{T} \log \left(\sum_{j=1}^{k} \lambda_j \ell_{jt} \right).$$

In doing so, the accessible parameter space is being shrunk in the sense that the MixN-GARCH model with constant mixing weights becomes the linchpin of the new model, i.e., that when maximizing the likelihood, the new model always nests the *optimal* one with constant weights for $\gamma = 0$. This restriction basically avoids the interaction of component and mixing parameters, except for the *leverage related* γ, and dramatically improves the out-of-sample quality as shown in Table 2.7 in comparison with the second variant, MixN-GARCH-LIKW, for which all parameters are estimated jointly.

Practically, we estimate $\boldsymbol{\nu}$ ceteris paribus using a *reduced EM* (REM) algorithm derived from the standard EM for mixtures of normals, so that $\boldsymbol{\nu}$ can be used in-place in a nested optimization, eluding a two-step procedure. Let $\boldsymbol{\ell}$ be the $k \times T$ matrix of the component-wise likelihood values (given the current estimate of the component models from the outer estimation of the GARCH parameters and γ). The REM algorithm cuts off the estimation of the component-wise density parameters by leaving them constant and estimates the mixing weights only, i.e., it iterates over

$$\nu_{j,n+1} = \frac{1}{T} \sum_{t=1}^{T} \frac{\nu_{j,n} \ell_{jt}}{\sum_i \nu_{i,n} \ell_{it}},$$

where $\nu_{j,0} = 1/k$ is the initial starting value, and stops if $|\nu_{j,n+1} - \nu_{j,n}| \leq \epsilon$ with, e.g., $\epsilon = 10^{-4}$. Most notably, REM exhibits a linear rate of convergence and has been observed to slow down the nested estimation for $k \geq 3$. It nevertheless remained feasible in all our empirical testing. For preventing the degeneracy of mixture components in the outer estimation, we employ the extended augmented maximum likelihood estimator (EALE) introduced in Broda et al. (2013).

2.3.3 Asymmetric Mixed Normal GARCH

In order to capture the leverage effect, Alexander and Lazar (2009) propose two asymmetric extensions of the MixN-GARCH model defined by (2.1) and (2.2). As we will consider these in our empirical applications below, we introduce them here. The first of these extensions, MixN-GARCH-ASYM, uses the asymmetric GARCH specification of Engle (1990) i.e., the GARCH process driving the variance of mixture component j is given by

$$\sigma_{jt}^2 = \omega_j + \alpha_j(\varepsilon_{t-1} - \theta_j)^2 + \beta_j \sigma_{j,t-1}^2, \quad j = 1, \ldots, k, \tag{2.23}$$

where the θ_j's are the parameters monitoring the component-specific leverage effect. In particular, if $\theta_j > 0$, a negative shock will increase the next period's σ_{jt}^2 more than a positive shock; a multivariate version of MixN-GARCH-ASYM has been investigated in Haas et al. (2009). The second variant, MixN-GARCH-GJR, employs the model of Glosten et al. (1993), widely known as GJR–GARCH, and specifies the variance process of component j as

$$\sigma_{jt}^2 = \omega_j + \alpha_j \varepsilon_{t-1}^2 + \theta_j d_{t-1}^- \varepsilon_{t-1}^2 + \beta_j \sigma_{j,t-1}^2, \quad j = 1, \ldots, k,$$

where $d_{t-1}^- = 1$ if $\varepsilon_{t-1} < 0$ and $d_{t-1}^- = 0$ otherwise. As in (2.23), a positive θ_j implies that σ_{jt}^2 reacts more intensely to negative shocks than to positive shocks.

2.4 Empirical Results

The empirical analysis is based on the major equity indices DAX30, S&P500, DJIA30, NIKKEI225 and NASDAQ COMPOSITE (10 years of data, dating back from July 7th, 2009) as well as the exchange rates JPY/EUR and USD/EUR (5 years of data, dating back from July 7th, 2009). All results (in-sample and out-of-sample) are based on daily percentage log returns, $\varepsilon_t = 100\,(\log P_t - \log P_{t-1})$, where P_t is the daily closing level of the index at time t.

As discussed, in this paper, we propose the two new models MixN-GARCH-LIK and MixN-GARCH-LOG, and their modeling and forecasting properties are described in this section. As a brief summary, of the two models it is MixN-GARCH-LIK that outperforms all its competitors by quite a huge margin. In fact, in the many out-of-sample forecasting exercises we discuss below it is this model, MixN-GARCH-LIK, that, for most summary statistics, archives the best scores independent of the datasets considered. For simplicity, we only entertain one and two component models in this paper, but results are also heavily in favor for MixN-GARCH-LIK when comparing three component models.[12]

[12]The results are available from the authors on request.

2.4.1 In-Sample Fit

For assessing in-sample properties, we fit all models under study to the entire data range, i.e., in Table 2.1 and 2.2 we show the likelihood values and BIC measures of all models and data sets. We focus on the BIC because the literature on mixture models provides some theoretical and empirical justification for its appropriateness and good performance, in particular for selecting the number of mixture components (see, e.g., Keribin, 2000; Francq et al., 2001; and Frühwirth-Schnatter, 2010, Ch. 4).

As expected, the pure likelihood values favor the two component models and center around the MixN-GARCH-ASYM and MixN-GARCH-GJR (both models with 11 free parameters). What is (perhaps) surprising is the fact that BIC, which favors less densely parameterized models, also has an overall tendency towards the two component models. In fact, for all data sets, the BIC signals superiority of the two–component models and, of those, the MixN-GARCH-GJR model wins in three out of the seven cases, even though this model has the highest parametrization, with 11 free parameters. However, the BIC of the MixN-GARCH-GJR model is not far from the two newly proposed ones, MixN-GARCH-LIK and MixN-GARCH-LOG, and as mentioned before, it is these models that shine above all in the more recognized out-of-sample forecasting comparison.

2.4.2 Forecasting Performance

More flexible models (e.g., all types of two component models) should be expected to provide an excellent in-sample fit to virtually any return series compared with more traditional (one component) GARCH-type models including the ones that can model several asymmetries, but the concern remains as to whether the additional parametrization and the nontrivial computational aspects of the feedback between different components warrant its use. To judge this, we compare the empirical performance of the one-step-ahead predictive cdfs across models using tests for uniformity (see below) as well as a variety of tests concentrating on the left tail of the return distribution as in Broda et al. (2013).

For all models considered, we re-estimate the model parameters every 20 trading days (about once a month), so that each estimation contains 2% of new data. Our analysis is based on the realized predictive cdf values obtained from evaluating the one-step-ahead cdf forecasts at the realized returns. If the model is correct, it is well-known that these are independently and uniformly distributed over the unit interval (Rosenblatt, 1952).

Let $\hat{p}_t = \hat{F}_{t-1}(\varepsilon_t; \hat{\boldsymbol{\theta}}_{t-h})$, $t = 1, \ldots, N$, be the sequence of realized predictive cdf values, noting that, for each t, the parameter vector is estimated using information (in this case, just the past returns) up to and including time $t - h$, where h is a value in $\{1, 2, \ldots, 20\}$, but the entire return series up to time $t - 1$

is used in the model filter. Finally, this predictive cdf is evaluated at the actual return at time t. Denote the collection of these N values as vector $\hat{\boldsymbol{p}}$. Further let $\hat{\boldsymbol{p}}^{[s]}$ denote the sorted vector, $\hat{p}_1^{[s]} \leq \hat{p}_2^{[s]} \leq \cdots \leq \hat{p}_N^{[s]}$. The Anderson-Darling (AD) and Cramér-von Mises (CM) test statistics are given respectively by

$$\text{AD} = -N - \sum_{i=1}^{N} \frac{2i-1}{N} \left(\log(\hat{p}_i^{[s]}) + \log(1 - \hat{p}_{N-i+1}^{[s]}) \right)$$

and

$$\text{CM} = \frac{1}{12N} + \sum_{i=1}^{N} \left(\frac{2i-1}{2N} - \hat{p}_i^{[s]} \right)^2 .$$

In addition, we provide test statistics for the Kolmogorov-Smirnov (KS) test for uniformity, as well as the Jarque-Bera (JB) and Shapiro-Wilk (SW) tests for normality. It is important to note that we are testing the prediction quality over the whole support of the distribution, and not just the left tail (as we do below, for directly testing the quality of value at risk predictions). Table 2.3 shows the results. The statistics AD, CM and KS reveal the astonishing performance of model MixN-GARCH-LIK in comparison to its competitors: in four out of seven cases for AD and in three out of seven cases for both CM and KS, it is MixN-GARCH-LIK that scores highest. For JB and SW there is also a clear tendency towards the two component models but no obvious pattern towards a particular type arises.

We also consider VaR measures dedicated to the left tail, as these are of even greater interest from a risk management perspective. Table 2.4 shows the empirical coverage probabilities (as percentages) for the 1% and 5% VaR levels. The results in Table 2.4 confirm the superiority of the proposed models in four out of seven cases at the 1% level and three cases at the 5% level.

For further investigations of the VaR prediction quality, we adopt a simple quality measure based on the coverage error over the VaR levels up to $100\lambda\%$, see Kuester et al. (2006). The measure calculates the deviation between predictive cdf and uniform cdf and, thus, captures the excess of percentage violations over the VaR levels, where the deviation is defined as $100(F_U - \hat{F}_e)$ with F_U being the cdf of the standard uniform random variable and $\hat{F}_e$ referring to the empirical cdf formed from $\hat{\boldsymbol{p}}$. Building upon this metric we report the integrated root mean squared error (IRMSE) over the left tail up to the maximal VaR level of interest. The IRMSE employed herein is closely related to the CM statistic but with the sum truncated at $h = \lceil \lambda N \rceil$, i.e.,

$$\text{IRMSE} = \sqrt{\frac{1}{h} \sum_{i=1}^{h} \left(100 \frac{2i-1}{2N} - 100 \hat{p}_i^{[s]} \right)^2} .$$

The results on the IRMSE in Table 2.5 are also in line with our general observation that MixN-GARCH-LIK is the overall winning model.

Finally as in Broda et al. (2013), we also investigate the hit sequence of realized predictive VaR violations,

$$v_t = \mathbb{1}_{\varepsilon_t \le \hat{q}_t}, \quad \hat{q}_t = \widehat{\text{VaR}}_{t|t-1}(\lambda),$$

where $\mathbb{1}$ is the indicator function. Under the null of correct conditional coverage, the v_i are iid Bernoulli(λ). From this sequence, the test statistic $\text{LR}_{\text{CC}} = \text{LR}_{\text{UC}} + \text{LR}_{\text{IND}}$ is computed, as proposed in Christoffersen (1998), where LR_{UC} and LR_{IND} test for unconditional coverage and independence, respectively. As can be seen from Table 2.6 for the 1% VaR level, as well as for the 5% VaR level, for all tests a clear tendency arises toward MixN-GARCH-LIK. With only one exception MixN-GARCH-LIK is the best performing model for every test.

2.5 Conclusions and Further Extensions

In this paper, we relax the assumption of constant weights in the class of mixed normal GARCH processes and introduce two different flexible time–varying weight model structures. Current mixing weights (and hence the implied overall volatility) are either directly related to past innovations by logistic response functions or indirectly via their lagged component-specific likelihood contributions. In particular, the second model type allows non-linear feedback between its likelihood components, and so induces news impact curves with (partial) leverage effects. As demonstrated, this latter model delivers clear-cut superior out-of-sample performance compared to all entertained models; and this, over a variety of data sets. Important open issues to be addressed in future research include establishing the stationarity conditions for the model and the asymptotic properties of the (augmented) maximum likelihood estimator.

The model classes are quite rich, and future applications should entertain other choices of the parameters and form structures. As mentioned above, Engle and Ng (1993) show that the older the news, the smaller the impact on current and future volatility. Also for the leverage effect, it is well known from Bouchaud et al. (2001) that its decay time differs across assets, with stocks requiring about 10 days, and indices about 50 days. These authors also show that the serial correlation function describing the magnitude of the leverage effect in terms of lags can be fit with a (single) exponential, which can be directly related to the dynamics between current mixing weights and past model innovations.

In addition, more flexible and more asymmetric model structures might be useful in order to further account for the "down–market effect" or "panic effect", i.e., a "one–sided" leverage effect related to falling stock prices. In fact, according to Figlewski and Wang (2000), a rise in the stock price does not affect volatility at all. They find the leverage effect is just a "down-market effect" not being existent for positive news surprises. This can easily be incorporated in our models by extending the constant weight assumption just for negative innovations, and/or using non-parametric response functions. Moreover, in addition to, or instead of, relating current mixing weights to the past innovations and likelihood contributions, it might be advantageous to consider use of the conditional variance, skewness or kurtosis. Further improvements to forecasting performance could also be gained by use of weighted likelihood; see Paolella and Steude (2008).

Finally, extensions into a multivariate framework are possible. For example, a straightforward generalization of MixN-GARCH-LIK is derived, e.g., by using the multivariate mixture GARCH model in Haas et al. (2009), given that the process of the mixture weights in (2.21) is entirely likelihood driven and, hence, generally applicable to univariate as well as multivariate models. Alternatively, it appears possible to augment the EM algorithm approach used for the multivariate mixture-based GARCH model in Paolella and Polak (2013) to the model structure used herein, thus rendering estimation in high dimensions feasible. Another approach which is also feasible in high dimensions is the use of Independent Component Analysis methods. Given the tractability of the moment generating and characteristic function of the conditional mixed normal forecast distribution used in this paper, the methodology in Broda and Paolella (2009) and Broda et al. (2013) is directly applicable. These ideas are currently being pursued.

Appendix

A Figures and Tables

Figures and tables are provided on following pages.

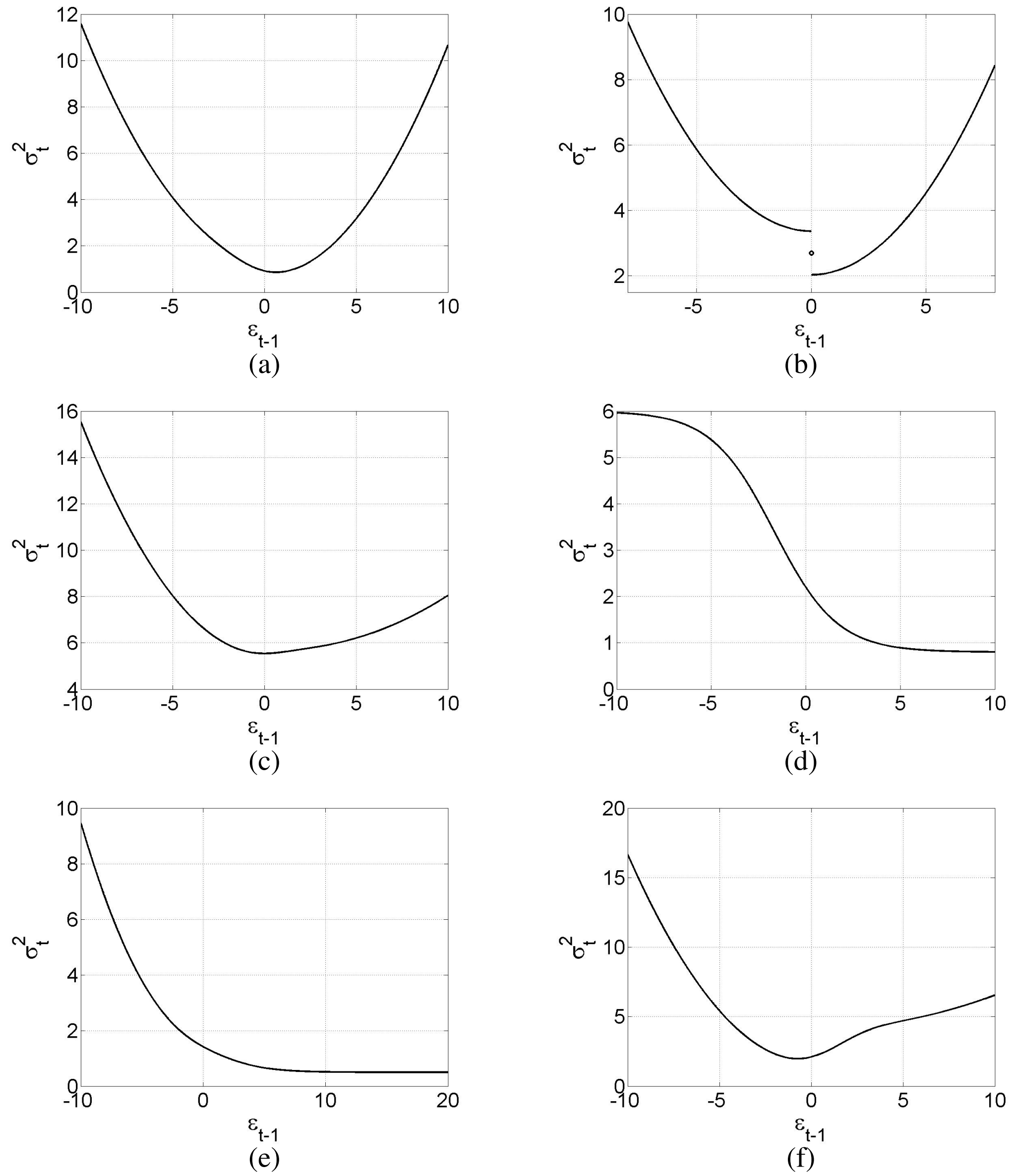


Figure 2.1: Shown are various possible shapes of the news impact curve (NIC; x-axis ε_{t-1}, y-axis σ_t^2) implied by the MixN-GARCH model (2.9) with time–varying weights described by (2.8). Panel (a) shows the NIC for the model with switching intercept (2.10) with $\gamma_0 = 1, \gamma_1 = 1, \omega_1 = 0.01, \omega_2 = 0.15, \alpha = 0.1$, and $\beta = 0.85$. Panel (b) illustrates the limiting case of this model with $\gamma_0 = 0, \gamma_1 = \infty, \omega_1 = 0.05, \omega_2 = 0.25, \alpha = 0.1$, and $\beta = 0.85$. Panel (c) shows the NIC for model (2.14) and $\gamma_0 = -1, \gamma_1 = 1, \omega = 0.02, \alpha_1 = 0.03, \alpha_2 = 0.1$, and $\beta = 0.9$. Panel (d) shows the NICs for the simple Gaussian mixture (constant variances) with $\gamma_0 = 1, \gamma_1 = 0.6, \omega_1 = 0.8, \omega_2 = 6$, and $\alpha = \beta = 0$. Panel (e) displays the NIC for the partial model (2.19) with $\gamma_0 = -0.5, \gamma_1 = 0.7, \omega_1 = 0.5, \omega_2 = 0.1, \alpha = 0.08$, and $\beta = 0.9$, and Panel (f) pertains to the general diagonal specification (2.17) with $\gamma_0 = -2, \gamma_1 = 1, \omega_1 = 0.35, \omega_2 = 0.05, \alpha_1 = 0.03, \alpha_2 = 0.15, \beta_1 = 0.9$, and $\beta_2 = 0.8$.

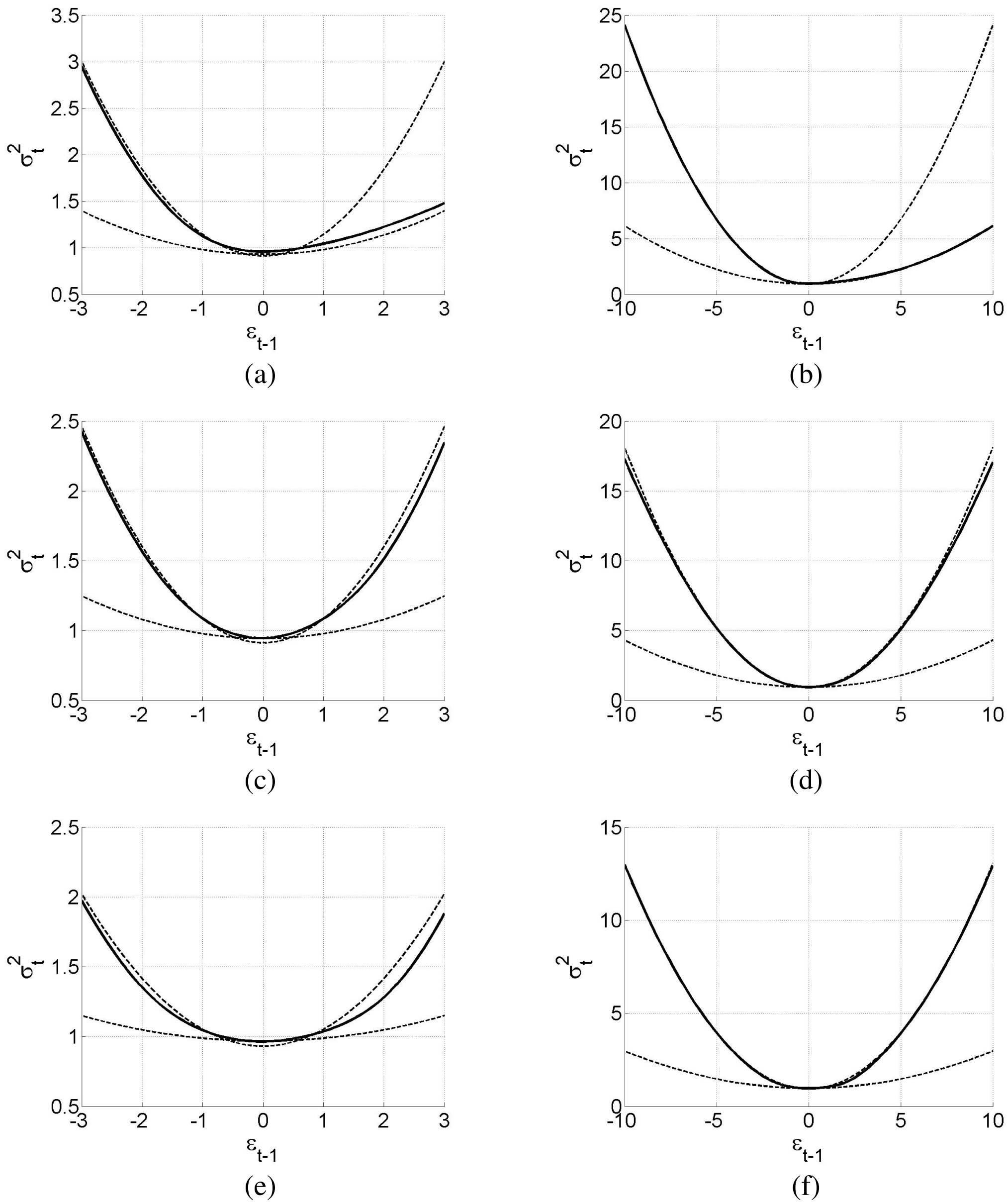


Figure 2.2: News impact curves (NIC) for selected estimates of MixN-GARCH-LIK and MixN-GARCH-LOG from the out-of-sample forecast exercise in Section 2.4. Figures (y-axis σ_t^2, x-axis ε_{t-1}) on the left show the magnified center of the figures on the right. Bold lines denote the NIC of the overall mixture, dashed lines denote the component-wise NICs. For MixN-GARCH-LIK, the leverage effect is particularly present in the range $-3 \leq \varepsilon_{t-1} \leq 3$, where for all data sets under study at least 90% of the (percentage log-) returns are located. Panel (a) and (b) are based on MixN-GARCH-LOG, the remaining panels on MixN-GARCH-LIK. Panel (a)–(d) show NICs for the DAX30 returns data as used in Table 2.3, whereas panel (e) and (f) use NASDAQ COMPOSITE data. The estimated parameter are $\gamma_0 = -0.21, \gamma_1 = -1.05, \mu_1 = -0.33, \mu_2 = 0.21, \omega_1 = 0.07, \omega_2 = 0.002, \alpha_1 = 0.23, \alpha_2 = 0.05, \beta_1 = 0.85, \beta_2 = 0.93$ for panel (a/b), $\gamma = 0.77, \mu_1 = 0.26, \mu_2 = -0.28, \omega_1 = 0.001, \omega_2 = 0.04, \alpha_1 = 0.03, \alpha_2 = 0.17, \beta_1 = 0.94, \beta_2 = 0.87$ for (c/d), and $\gamma = 7.02, \mu_1 = 0.22, \mu_2 = -0.16, \omega_1 = 0.001, \omega_2 = 0.03, \alpha_1 = 0.02, \alpha_2 = 0.12, \beta_1 = 0.97, \beta_2 = 0.9$ for (e/f).

model	free param.	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Normal-GARCH	4	-6089.67	-5254.12	-5143.05	-6334.56	-2372.40	-2220.49	-6387.07
Normal-ASYM-GARCH	5	-6050.43	-5193.92	-5086.38	-6285.86	-2370.69	-2220.22	-6357.60
Normal-GJR-GARCH	5	-6050.10	-5180.52	-5086.28	-6298.24	-2370.20	-2220.49	-6350.38
Normal-EGARCH	5	-6043.56	-5174.52	-5076.13	-6291.27	-2366.82	-2222.25	-6351.47
MixN-GARCH	9	-6036.59	-5190.73	-5076.31	-6277.75	-2332.27	-2212.91	-6355.07
MixN-GARCH-ASYM	11	-6010.95	-5140.66	-5035.04	**-6245.84**	-2328.32	**-2209.64**	-6324.46
MixN-GARCH-GJR	11	**-6003.13**	**-5122.94**	**-5028.65**	-6251.15	-2329.19	-2211.10	**-6309.50**
MixN-GARCH-LIK	10	-6031.16	-5188.23	-5073.16	-6277.30	-2330.67	-2209.86	-6349.75
MixN-GARCH-LOG	11	-6024.64	-5181.35	-5067.54	-6274.52	**-2326.73**	-2210.12	-6335.01

Table 2.1: In-sample likelihood values for all single- and multi-component mixture GARCH models and all data sets under study. In-sample statistics are based on complete data sets as used in Table 2.3. For comparison, all models include a location parameter for the density. Entries in boldface denote the best results per data set.

model	free param.	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Normal-GARCH	4	12212.11	10541.01	10318.87	12701.89	4775.77	**4471.95**	12806.90
Normal-ASYM-GARCH	5	12141.82	10428.79	10213.71	12612.68	4780.10	4479.16	12756.15
Normal-GJR-GARCH	5	12141.15	10402.00	10213.52	12637.44	4779.12	4479.69	12741.72
Normal-EGARCH	5	12128.07	10390.00	10193.21	12623.50	4772.36	4483.21	12743.89
MixN-GARCH	9	12138.71	10446.99	10218.16	12621.03	**4726.48**	4487.76	12775.67
MixN-GARCH-ASYM	11	12112.00	10371.41	10160.19	**12581.78**	4741.80	4504.44	12739.02
MixN-GARCH-GJR	11	**12096.37**	**10335.98**	**10147.40**	12592.41	4743.54	4507.37	**12709.10**
MixN-GARCH-LIK	10	12136.05	10450.18	10220.04	12628.33	4731.03	4489.40	12773.23
MixN-GARCH-LOG	11	12131.18	10444.62	10217.00	12630.95	4730.88	4497.66	12751.93

Table 2.2: BIC values for all single- and multi-component mixture GARCH models and all data sets under study. In-sample statistics are based on complete data sets as used in Table 2.3 For comparison, all models include a location parameter for the density. Entries in boldface denote the best results per data set.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Anderson-Darling							
Normal-GARCH	4.24***	3.70**	3.37**	3.29**	7.03***	3.37**	2.86**
Normal-ASYM-GARCH	3.94***	3.16**	2.75**	2.44*	6.79***	3.14**	2.73**
Normal-GJR-GARCH	4.01***	2.75**	2.47*	2.84**	6.86***	3.35**	2.70**
Normal-EGARCH	4.66***	3.17**	2.46*	2.70**	6.41***	2.91**	2.77**
MixN-GARCH	1.01	0.91	1.23	1.16	**0.65**	0.62	1.53
MixN-GARCH-ASYM	1.18	1.15	1.12	1.42	0.75	1.16	1.37
MixN-GARCH-GJR	1.12	1.31	1.18	1.20	0.69	1.04	**1.20**
MixN-GARCH-LIK	**0.84**	**0.57**	**0.95**	**1.11**	0.66	**0.61**	1.26
MixN-GARCH-LOG	2.67**	2.16*	2.08*	1.25	0.83	1.07	3.08**
Cramér-von Mises							
Normal-GARCH	0.81***	0.64**	0.61**	0.57**	1.17***	0.69**	0.54**
Normal-ASYM-GARCH	0.73**	0.51**	0.47**	0.37*	1.14***	0.65**	0.54**
Normal-GJR-GARCH	0.72**	0.45*	0.43*	0.47**	1.15***	0.69**	0.51**
Normal-EGARCH	0.72**	0.37*	0.32	0.44*	1.07***	0.61**	0.50**
MixN-GARCH	0.13	0.10	0.18	0.19	**0.07**	0.10	0.18
MixN-GARCH-ASYM	0.14	0.15	0.15	0.20	0.07	0.20	0.21
MixN-GARCH-GJR	0.13	0.19	0.15	0.17	0.07	0.18	0.18
MixN-GARCH-LIK	**0.11**	**0.07**	**0.14**	0.19	0.07	**0.10**	**0.15**
MixN-GARCH-LOG	0.32	0.25	0.30	**0.14**	0.07	0.19	0.38*
Kolmogorov-Smirnov (test statistics are scaled up by factor 100)							
Normal-GARCH	3.93***	3.32**	3.12**	3.01**	6.32***	5.05***	3.41***
Normal-ASYM-GARCH	3.60***	2.79*	2.84*	2.53	6.39***	4.99***	3.32**
Normal-GJR-GARCH	3.70***	2.96**	2.88*	2.90**	6.34***	5.05***	3.26**
Normal-EGARCH	3.22**	2.27	2.54	2.72*	6.04***	4.97***	3.09**
MixN-GARCH	1.77	1.49	2.08	1.89	**1.69**	**2.30**	1.97
MixN-GARCH-ASYM	2.07	1.45	1.92	2.18	1.94	3.07	**1.82**
MixN-GARCH-GJR	1.78	1.78	1.79	2.19	1.97	3.02	1.85
MixN-GARCH-LIK	**1.65**	**1.23**	**1.76**	1.90	1.76	2.37	1.97
MixN-GARCH-LOG	2.32	2.58	2.72*	**1.56**	1.91	2.96	2.43
Jarque-Bera							
Normal-GARCH	243.98***	272.92***	316.41***	163.89***	310.80***	41.75***	89.09***
Normal-ASYM-GARCH	106.97***	208.96***	164.07***	119.16***	338.56***	43.00***	134.64***
Normal-GJR-GARCH	109.15***	319.48***	265.59***	149.15***	312.85***	45.41***	114.31***
Normal-EGARCH	102.35***	320.04***	256.53***	114.18***	274.67***	48.70***	155.01***
MixN-GARCH	38.35***	**38.57*****	35.79***	4.45	**15.57*****	4.54	35.31***
MixN-GARCH-ASYM	**11.14*****	40.72***	**26.33*****	15.16***	51.59***	10.08**	48.45***
MixN-GARCH-GJR	23.04***	64.59***	38.78***	12.22***	30.67***	8.53**	43.99***
MixN-GARCH-LIK	25.57***	40.92***	33.76***	**3.76**	15.94***	**4.48**	**17.99*****
MixN-GARCH-LOG	57.72***	68.96***	72.89***	27.93***	22.68***	21.38***	37.40***
Shapiro-Wilk (test statistic ν is transformed by $1000(1-\nu)$)							
Normal-GARCH	10.24***	11.30***	12.14***	8.80***	29.83***	6.18***	6.12***
Normal-ASYM-GARCH	7.55***	9.72***	8.70***	7.12***	30.37***	6.42***	7.56***
Normal-GJR-GARCH	7.89***	11.55***	10.64***	8.18***	29.67***	6.51***	6.96***
Normal-EGARCH	7.62***	12.28***	10.84***	6.99***	28.42***	6.70***	8.23***
MixN-GARCH	3.52***	**3.19*****	3.36***	1.85***	**4.26*****	1.31	4.03***
MixN-GARCH-ASYM	**2.16*****	3.37***	**2.73*****	2.42***	8.27***	2.36	4.23***
MixN-GARCH-GJR	2.70***	4.40***	3.23***	2.24***	6.38***	2.11	4.17***
MixN-GARCH-LIK	2.93***	3.34***	3.31***	**1.72****	4.33***	**1.29**	**2.96*****
MixN-GARCH-LOG	4.22***	4.36***	4.66***	3.37***	5.62***	3.67***	3.60***

Table 2.3: Anderson-Darling, Cramér-von Mises, Kolmogorov-Smirnov, Jarque-Bera and Shapiro-Wilk test statistics for all models and data sets under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. All values are based on evaluating the one-step-ahead out-of-sample distribution forecasts at the observed return data, estimated throughout a rolling window of $1,000$ trading days, covering 10 years of equity returns and 5 years of FX returns, dating back from July 7, 2009. For $k > 2$, the model parameters are updated every 20 days, while for single-component models, we update in every step as, otherwise, EGARCH would suffer from non-stationary issues preventing a comparison.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
Normal-GARCH	1.59***	1.77***	1.68***	1.49**	2.48***	1.10	1.18
Normal-ASYM-GARCH	1.71***	1.89***	1.83***	1.88***	2.38***	1.12	1.26
Normal-GJR-GARCH	1.62***	1.76***	1.60***	1.65***	2.40***	1.09	1.29*
Normal-EGARCH	1.81***	1.93***	2.06***	1.86***	2.71***	1.25	1.41**
MixN-GARCH	1.09	1.24	1.21	0.87	1.64**	1.09	**1.00**
MixN-GARCH-ASYM	1.07	1.48**	1.60***	1.29*	1.65**	**1.03**	1.11
MixN-GARCH-GJR	1.15	1.47**	1.32*	1.18	**1.52****	1.05	0.91
MixN-GARCH-LIK	**1.02**	**1.16**	**1.16**	0.78	1.68**	1.09	0.95
MixN-GARCH-LOG	1.31*	1.50**	1.41**	**1.12**	1.87***	1.14	1.03
5% VaR							
Normal-GARCH	5.98**	5.62*	5.45	**5.57**	5.74	4.68	5.94**
Normal-ASYM-GARCH	6.44***	5.57	5.19	5.82**	5.68	4.86	5.79**
Normal-GJR-GARCH	6.34***	5.64*	**5.19**	5.64*	5.66	4.68	5.81**
Normal-EGARCH	6.93***	6.35***	5.61*	5.59*	5.78	5.29	6.05***
MixN-GARCH	5.75**	5.39	5.50	5.77**	5.66	5.00	5.91**
MixN-GARCH-ASYM	6.10***	5.68*	5.45	6.01**	**5.19**	4.72	5.62*
MixN-GARCH-GJR	5.59*	5.91**	5.44	5.80**	5.52	5.05	**5.27**
MixN-GARCH-LIK	**5.37**	**4.84**	5.25	5.72*	5.66	5.03	5.48
MixN-GARCH-LOG	6.30***	5.93**	5.76**	6.08***	5.46	**5.00**	6.64***

Table 2.4: Predicted VaR coverage percentages (point estimates) at the 1% and 5% level for all models under study. Entries in boldface denote the best (closest to the true value) estimate. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 2.3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
Normal-GARCH	0.32	0.38	0.39	0.30	0.52	0.24	0.23
Normal-ASYM-GARCH	0.33	0.43	0.43	0.35	0.52	0.33	0.26
Normal-GJR-GARCH	0.39	0.35	0.41	0.38	0.52	0.25	0.23
Normal-EGARCH	0.45	0.47	0.48	0.35	0.53	0.27	0.29
MixN-GARCH	**0.05**	0.21	0.18	0.12	0.31	0.13	0.08
MixN-GARCH-ASYM	0.10	0.25	0.26	0.21	0.31	**0.10**	0.11
MixN-GARCH-GJR	0.20	**0.14**	0.20	**0.08**	**0.24**	0.18	0.08
MixN-GARCH-LIK	0.05	0.18	**0.16**	0.13	0.31	0.13	**0.06**
MixN-GARCH-LOG	0.21	0.29	0.28	0.19	0.34	0.22	0.12
5% VaR							
Normal-GARCH	0.78	0.80	0.68	0.58	1.17	0.28	0.49
Normal-ASYM-GARCH	0.77	0.83	0.64	0.75	1.19	0.22	0.47
Normal-GJR-GARCH	0.69	0.81	0.69	0.67	1.22	0.26	0.50
Normal-EGARCH	1.02	1.08	0.82	0.70	1.35	0.22	0.68
MixN-GARCH	0.35	0.39	0.44	0.38	0.53	0.19	0.37
MixN-GARCH-ASYM	0.39	0.54	0.53	0.68	0.41	0.18	0.26
MixN-GARCH-GJR	0.29	0.47	0.43	0.53	**0.41**	0.16	**0.10**
MixN-GARCH-LIK	**0.14**	**0.13**	**0.22**	**0.29**	0.53	0.20	0.21
MixN-GARCH-LOG	0.64	0.61	0.61	0.60	0.66	**0.16**	0.53

Table 2.5: Integrated root mean squared error of the VaR prediction up to the 1% and 5% level for all models under study. Entries in boldface denote the best estimate. Results are based on the same out-of-sample exercise as in Table 2.3.

	level	model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Unconditional Coverage, LR_{UC}	1%	Normal-GARCH	7.33***	12.51***	10.30***	5.60**	19.81***	0.07	0.88
		Normal-ASYM-GARCH	11.38***	16.15***	14.89***	16.15***	18.02***	0.28	1.71
		Normal-GJR-GARCH	8.27***	12.51***	8.27***	9.26***	18.02***	0.07	2.21
		Normal-EGARCH	13.68***	17.45***	23.05***	14.89***	25.57***	0.63	4.08**
		MixN-GARCH	0.14	1.26	1.26	0.39	4.14**	0.07	**3.1e-4**
		MixN-GARCH-ASYM	0.14	5.60**	8.27***	2.21	5.16**	**1.2e-4**	0.32
		MixN-GARCH-GJR	0.56	4.81**	2.21	0.88	**3.23***	0.07	0.17
		MixN-GARCH-LIK	**0.03**	**0.56**	**0.56**	1.56	5.16**	0.07	0.05
		MixN-GARCH-LOG	2.21	5.60**	4.08**	**0.32**	7.45***	0.28	0.03
	5%	Normal-GARCH	4.97**	2.13	1.05	**1.65**	1.48	0.29	4.60**
		Normal-ASYM-GARCH	10.47***	1.65	0.25	3.57*	1.20	0.08	3.25*
		Normal-GJR-GARCH	8.92***	2.13	**0.17**	2.13	1.20	0.29	3.57*
		Normal-EGARCH	18.50***	9.42***	1.88	1.88	1.48	0.23	5.75**
		MixN-GARCH	2.95*	0.88	1.44	2.95*	1.20	**6.5e-4**	4.24**
		MixN-GARCH-ASYM	6.17**	2.39	1.05	5.36**	**0.12**	0.17	2.13
		MixN-GARCH-GJR	1.88	4.24**	1.05	3.25*	0.72	0.01	**0.34**
		MixN-GARCH-LIK	**0.72**	**0.16**	0.34	2.66	1.20	0.01	1.23
		MixN-GARCH-LOG	8.43***	4.60**	2.95*	6.17**	0.53	**6.5e-4**	13.31***
Independence, LR_{IND}	1%	Normal-GARCH	0.21	0.08	1.54	**0.28**	3.98**	0.33	0.77
		Normal-ASYM-GARCH	0.10	**0.04**	1.84	1.91	4.29**	0.37	0.87
		Normal-GJR-GARCH	0.17	0.08	1.41	1.48	4.29**	0.33	0.92
		Normal-EGARCH	**0.06**	1.99	2.33	1.84	1.02	0.42	1.09
		MixN-GARCH	1.06	0.82	0.82	0.43	0.92	0.33	0.54
		MixN-GARCH-ASYM	0.63	1.21	**0.17**	0.92	0.79	**0.28**	0.67
		MixN-GARCH-GJR	0.72	1.15	0.92	0.77	0.65	0.33	**0.46**
		MixN-GARCH-LIK	1.17	0.72	0.72	0.32	0.79	0.33	0.50
		MixN-GARCH-LOG	0.56	1.21	1.09	0.67	**0.58**	0.37	0.59
	5%	Normal-GARCH	4.40**	0.13	0.33	**0.11**	**0.79**	0.55	0.31
		Normal-ASYM-GARCH	4.50**	2.88*	3.39*	3.81*	1.90	0.40	1.20
		Normal-GJR-GARCH	**2.62**	0.18	0.70	1.80	1.90	0.55	0.28
		Normal-EGARCH	3.46*	2.75*	**0.12**	3.00*	1.73	2.88*	1.84
		MixN-GARCH	3.53*	0.13	1.30	0.17	1.90	0.29	0.22
		MixN-GARCH-ASYM	4.87**	5.16**	2.51	4.56**	1.70	0.47	1.80
		MixN-GARCH-GJR	7.39***	6.18**	2.51	3.67*	1.13	**0.24**	0.34
		MixN-GARCH-LIK	3.94**	**0.10**	0.58	0.15	1.90	**0.24**	**0.12**
		MixN-GARCH-LOG	3.89**	4.25**	0.17	1.05	1.26	0.29	0.36
Conditional Coverage, LR_{CC}	1%	Normal-GARCH	7.54**	12.58***	11.84***	5.89	23.80***	0.40	1.65
		Normal-ASYM-GARCH	11.47***	16.19***	16.73***	18.06***	22.31***	0.66	2.58
		Normal-GJR-GARCH	8.44**	12.58***	9.68**	10.74***	22.31***	0.40	3.14
		Normal-EGARCH	13.74***	19.44***	25.37***	16.73***	26.59***	1.06	5.17
		MixN-GARCH	1.20	2.08	2.08	**0.81**	5.06	0.40	**0.54**
		MixN-GARCH-ASYM	**0.77**	6.82*	8.44**	3.14	5.95	**0.28**	0.99
		MixN-GARCH-GJR	1.29	5.97	3.14	1.65	**3.88**	0.40	0.64
		MixN-GARCH-LIK	1.20	**1.29**	**1.29**	1.89	5.95	0.40	0.55
		MixN-GARCH-LOG	2.77	6.82*	5.17	0.99	8.04**	0.66	0.62
	5%	Normal-GARCH	9.38**	2.25	1.38	**1.77**	2.27	0.84	4.91
		Normal-ASYM-GARCH	14.97***	4.53	3.64	7.38**	3.10	0.48	4.45
		Normal-GJR-GARCH	11.54***	2.31	**0.87**	3.92	3.10	0.84	3.85
		Normal-EGARCH	21.95***	12.17***	2.00	4.88	3.21	3.11	7.59**
		MixN-GARCH	6.48*	1.01	2.73	3.12	3.10	0.29	4.46
		MixN-GARCH-ASYM	11.04***	7.55**	3.56	9.91**	1.83	0.64	3.92
		MixN-GARCH-GJR	9.27**	10.42**	3.56	6.92*	1.85	**0.25**	**0.68**
		MixN-GARCH-LIK	**4.66**	**0.26**	0.92	2.81	3.10	**0.25**	1.35
		MixN-GARCH-LOG	12.32***	8.85**	3.12	7.22*	**1.78**	0.29	13.67***

Table 2.6: Test statistics at the 1%- and 5%-VaR level, $LR_{CC} = LR_{UC} + LR_{IND}$, as described in Christoffersen (1998) for all models under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 2.3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Likelihood							
MixN-GARCH-LIK	-6031.16	-5188.23	-5073.16	-6277.30	-2330.67	**-2209.86**	-6349.75
MixN-GARCH-LIKW	**-6027.02**	**-5180.88**	**-5067.12**	**-6276.77**	**-2330.67**	-2210.78	**-6347.09**
tab:bic							
MixN-GARCH-LIK	**12136.05**	**10450.18**	**10220.04**	**12628.33**	**4731.03**	**4489.40**	**12773.23**
MixN-GARCH-LIKW	12144.14	10451.86	10224.33	12643.65	4746.51	4506.72	12784.28
Anderson-Darling							
MixN-GARCH-LIK	**0.84**	**0.57**	**0.95**	1.11	**0.66**	**0.61**	**1.26**
MixN-GARCH-LIKW	1.00	0.97	1.22	**1.11**	0.66	0.61	1.32
Cramér-von Mises							
MixN-GARCH-LIK	**0.11**	**0.07**	**0.14**	0.19	**0.07**	**0.10**	**0.15**
MixN-GARCH-LIKW	0.13	0.11	0.17	**0.17**	0.07	0.10	0.15
Kolmogorov-Smirnov (test statistics are scaled up by factor 100)							
MixN-GARCH-LIK	**1.65**	**1.23**	**1.76**	1.90	**1.76**	2.37	1.97
MixN-GARCH-LIKW	1.71	1.43	1.87	**1.81**	1.82	**2.35**	**1.84**
Ljung-Box ($m = 20$ lags)							
MixN-GARCH-LIK	28.26	**35.72****	**32.71***	**17.45**	18.65	**18.94**	**32.09***
MixN-GARCH-LIKW	**27.54**	36.25**	33.15*	17.60	**18.59**	18.98	33.70*
Jarque-Bera							
MixN-GARCH-LIK	25.57***	**40.92*****	**33.76*****	**3.76**	**15.94*****	**4.48**	**17.99*****
MixN-GARCH-LIKW	**20.90*****	84.62***	36.05***	3.96	16.00***	4.50	30.45***
Shapiro-Wilk (test statistic ν is transformed by $1000(1-\nu)$)							
MixN-GARCH-LIK	2.93***	**3.34*****	**3.31*****	1.72**	**4.33*****	**1.29**	**2.96*****
MixN-GARCH-LIKW	**2.60*****	5.15***	3.41***	**1.69****	4.34***	1.29	3.93***
1% VaR							
MixN-GARCH-LIK	**1.02**	**1.16**	**1.16**	0.78	1.68**	**1.09**	**0.95**
MixN-GARCH-LIKW	1.03	1.24	1.30*	**0.92**	**1.64****	1.09	0.89
5% VaR							
MixN-GARCH-LIK	**5.37**	**4.84**	**5.25**	**5.72***	**5.66**	5.03	5.48
MixN-GARCH-LIKW	5.66*	5.25	5.31	5.88**	5.72	**5.00**	**5.44**
RMSE up to 1% VaR							
MixN-GARCH-LIK	0.05	**0.18**	**0.16**	0.13	**0.31**	**0.13**	**0.06**
MixN-GARCH-LIKW	**0.04**	0.24	0.20	**0.08**	0.31	0.13	0.09
RMSE up to 5% VaR							
MixN-GARCH-LIK	**0.14**	**0.13**	**0.22**	**0.29**	**0.53**	**0.20**	**0.21**
MixN-GARCH-LIKW	0.20	0.33	0.36	0.39	0.54	0.20	0.25
Unconditional Coverage, LR_{UC}							
MixN-GARCH-LIK	**0.03**	**0.56**	**0.56**	1.56	5.16**	**0.07**	**0.05**
MixN-GARCH-LIKW	**0.03**	1.26	2.21	**0.17**	**4.14****	**0.07**	0.39
Independence, LR_{IND}							
MixN-GARCH-LIK	**1.17**	**0.72**	**0.72**	**0.32**	**0.79**	**0.33**	0.50
MixN-GARCH-LIKW	**1.17**	0.82	0.92	0.46	0.92	**0.33**	**0.43**
Conditional Coverage, LR_{CC}							
MixN-GARCH-LIK	**1.20**	**1.29**	**1.29**	1.89	5.95	**0.40**	**0.55**
MixN-GARCH-LIKW	**1.20**	2.08	3.14	**0.64**	**5.06**	**0.40**	0.81

Table 2.7: Results as in Table 2.1–2.6 but for MixN-GARCH-LIK and MixN-GARCH-LIKW. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 2.3.

Bibliography

Alexander, C. (2008). *Practical Financial Econometrics*. John Wiley & Sons, Chichester.

Alexander, C. and Lazar, E. (2006). Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling. *Journal of Applied Econometrics*, 21:307–336.

Alexander, C. and Lazar, E. (2009). Modelling Regime–specific Stock Price Volatility. *Oxford Bulletin of Economics and Statistics*, 71:761–797.

Aparicio, F. M. and Estrada, J. (2001). Empirical Distributions of Stock Returns: European Securities Markets, 1990–95. *European Journal of Finance*, 7:1–21.

Asai, M. and McAleer, M. (2011). Alternative Asymmetric Stochastic Volatility Models. *Econometric Reviews*, 30(5):548–564.

Bauwens, L., Preminger, A., and Rombouts, J. (2006). Regime Switching GARCH Models. CORE Discussion Paper 2006/11, Université Catholique de Louvain.

Bekaert, G. and Gray, S. F. (1998). Target Zones and Exchange Rates: An Empirical Investigation. *Journal of International Economics*, 45:1–35.

Black, F. (1976). Studies in Stock Price Volatility Changes. In *American Statistical Association, Proceedings of the Business and Economic Statistics Section*, pages 177–181.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bouchaud, J.-P., Matacz, A., and Potters, M. (2001). Leverage Effect in Financial Markets: The Retarded Volatility Model. *Physical Review Letters*, 87(22):1–4.

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013). Stable mixture GARCH models. *Journal of Econometrics*, 172(2):292–306.

Broda, S. A. and Paolella, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of Financial Econometrics*, 7(4):412–436.

Cheng, X., Yu, P. L. H., and Li, W. K. (2009). On a Dynamic Mixture GARCH Model. *Journal of Forecasting*, 28:247–265.

Christie, A. (1982). The Stochastic Behavior of Common Stock Variance: Value, Leverage, and Interest Rate Effects. *Journal of Financial Economics*, 10:407–432.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Engle, R. F. (1990). Stock Volatility and the Crash of ’87: Discussion. *Review of Financial Studies*, 3:103–106.

Engle, R. F. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48:1749–1778.

Fama, E. F. (1965). The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105.

Figlewski, S. and Wang, X. (2000). Is the “Leverage Effect” a Leverage Effect? Working paper series 00-37, New York University.

Fornari, F. and Mele, A. (1997). Sign- and Volatility-Switching ARCH Models: Theory and Applications to International Stock Markets. *Journal of Applied Econometrics*, 12:49–65.

Francq, C., Roussignol, M., and Zakoïan, J.-M. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22(2):197–220.

Frühwirth-Schnatter, S. (2010). *Finite Mixture and Markov Switching Models*. Springer.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993). Nonlinear Dynamic Structures. *Econometrica*, 61(4):871–907.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Haas, M., Mittnik, S., and Mizrach, B. (2006a). Assessing Central Bank Credibility During the EMS Crises: Comparing Option and Spot Market-Based Forecasts. *Journal of Financial Stability*, 2:28–54.

Haas, M., Mittnik, S., and Paolella, M. S. (2004a). A New Approach to Markov Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Haas, M., Mittnik, S., and Paolella, M. S. (2004b). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.

Haas, M., Mittnik, S., and Paolella, M. S. (2009). Asymmetric Multivariate Normal Mixture GARCH. *Computational Statistics and Data Analysis*, 53:2129–2154.

Haas, M., Mittnik, S., Paolella, M. S., and Steude, S. C. (2006b). Analyzing and Exploiting Asymmetries in the News Impact Curve. *National Centre of Competence in Research Financial Valuation and Risk Management, Working Paper Series*.

Haas, M. and Paolella, M. S. (2012). Mixture and Regime-switching GARCH Models. In Bauwens, L., Hafner, C. M., and Laurent, S., editors, *Handbook of volatility models and their applications*, number 3.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357–384.

Hasanhodzic, J. and Lo, A. W. (2011). Black's Leverage Effect is Not Due to Leverage. *SSRN eLibrary*.

Hens, T. and Steude, S. C. (2009). The leverage effect without leverage. *Finance Research Letters*, 6(2):83–94.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62(1):49–66.

Kim, T.-H. and White, H. (2004). On More Robust Estimation of Skewness and Kurtosis. *Finance Research Letters*, 1:56–73.

Klaster, M. A. and Knot, K. H. W. (2002). Toward an Econometric Target Zone Model with Endogenous Devaluation Risk for a Small Open Economy. *Economic Modelling*, 19:509–529.

Kon, S. J. (1984). Models of Stock Returns: A Comparison. *The Journal of Finance*, 39:147–165.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value–at–Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4:53–89.

Ling, S. and McAleer, M. (2002). Stationarity and the Existence of Moments of a Family of GARCH Processes. *Journal of Econometrics*, 106:109–117.

McAleer, M. and da Veiga, B. (2008). Forecasting value-at-risk with a parsimonious portfolio spillover GARCH (PS-GARCH) model. *Journal of Forecasting*, 27(1):1–19.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley.

Neely, C. J. (1999). Target Zones and Conditional Volatility: The Role of Realignments. *Journal of Empirical Finance*, 6:177–192.

Neftci, S. N. (2000). Value at Risk Calculations, Extreme Events, and Tail Estimation. *Journal of Derivatives*, 7:23–28.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59:347–37.

Palm, F. C. (1996). GARCH Models of Volatility. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume 14, pages 209–240. Elsevier Science, Amsterdam.

Paolella, M. S. and Polak, P. (2013). COMFORT-CCClass: A Common Market Factor Non-Gaussian Returns Model. Swiss Finance Institute (SFI) Research Paper Series.

Paolella, M. S. and Steude, S. C. (2008). Risk prediction: a DWARF-like approach. *The Journal of Risk Model Validation*, 2(1):25–43.

Park, B.-J. (2011). Asymmetric herding as a source of asymmetric return volatility. *Journal of Banking & Finance*, 35(10):2657–2665.

Rodriguez, M. J. and Ruiz, E. (2012). Revisiting Several Popular GARCH Models with Leverage Effect: Differences and Similarities. *Journal of Financial Econometrics*, 10:637–668.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23:470–472.

Tashman, A. and Frey, R. J. (2009). Modeling risk in arbitrage strategies using finite mixtures. *Quantitative Finance*, 9:495–503.

Tucker, A. L. and Pond, L. (1988). The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes. *Review of Economics and Statistics*, 70:638–647.

Vlaar, P. J. G. and Palm, F. C. (1993). The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps. *Journal of Business and Economic Statistics*, 11(3):351–360.

Wong, C. S. and Li, W. K. (2001). On a Logistic Mixture Autoregressive Model. *Biometrika*, 88:833–846.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.

Chapter 3

Portfolio optimization using the noncentral t distribution

Portfolio optimization using the noncentral t distribution*

Simon A. Broda[a,b] Jochen Krause[c] Marc S. Paolella[c,d]

[a]*Department of Economics and Econometrics, University of Amsterdam*

[b]*Tinbergen Institute Amsterdam*

[c]*Department of Banking and Finance, University of Zurich, Switzerland*

[d]*Swiss Finance Institute*

Abstract

A new multivariate GARCH model is proposed that combines the noncentral Student's t distribution with the dynamic conditional correlation (DCC) filter of Engle (2002). The model accounts for most stylized facts of asset returns, including time varying volatility and correlation, fat tails and asymmetry, as well as non-ellipticity. A three-step estimation procedure is devised, and a new approximation to the density of the multivariate noncentral t is introduced. Considering minimum expected shortfall portfolios, a link between the saddlepoint approximation of expected shortfall in Broda and Paolella (2010) and the approach to portfolio optimization in Rockafellar and Uryasev (2000) is established and shown to significantly reduce computation times. An out-of-sample forecasting exercise based on the 30 components of the Dow Jones Industrial Average Index confirms the superiority of the new model compared to the Gaussian DCC model in forecasting quality and portfolio performance.

Keywords — Expected Shortfall, Noncentral t distribution, Non-ellipticity; Portfolio Optimization; Saddlepoint Approximation; Transformed means.

*Part of the research of M. S. Paolella has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the *Swiss National Science Foundation*.

3.1 Introduction

In Broda and Paolella (2010) a new saddlepoint approximation for the expected shortfall (ES) of transformed means is introduced. The approximation proposed therein allows the evaluation of expected shortfall for random variables with an unknown or inaccessible ES expression but with a stochastic representation in form of random variables that possess a moment generating function. The result facilitates the fast, reliable and accurate evaluation of ES for many distributions of practical interest, e.g., Azzalini's skewed t distribution, and the noncentral Student's t distribution, for which otherwise numerical techniques have to be resorted to. We demonstrate the usefulness of the result in Broda and Paolella (2010) using the latter and develop an application in portfolio optimization, of which an empirical exercise is shown using the 30 components of the Dow Jones Industrial Average index. Being a location scale mixture of normals (see, e.g., Mencía and Sentana, 2009), the (singly) noncentral t distribution, hereafter MVNCT, can be represented by random variables with an easy moment generating function, and thus, lends itself for use with Broda and Paolella's result. The MVNCT shares the property of the Gaussian and the multivariate generalized hyperbolic (of which the multivariate t is a limiting case, but not the MVNCT), that weighted sums of the univariate marginals remain in the distributional class. That is, under the assumption that the multivariate set of asset returns at a particular point in time follows a MVNCT, the distribution of the portfolio return is a univariate noncentral t. In addition, it comes with a tail dependence that is less extreme compared to the one of the multivariate generalized hyperbolic, see Jondeau (2010, Section 3).

We develop a numerically fast method to determine the weights corresponding to the min-ES portfolio (MESP). The MESP is similar to the classic minimum-variance portfolio in standard portfolio theory but utilizes a more advanced left-tailed risk measure, the expected shortfall. This is relevant, and will differ from the choice based on the variance, in particular when the distribution of asset returns is non-elliptical, see Embrechts et al. (1999, Theorem 1). For modeling the volatility clustering and dynamic correlations, the dynamic conditional correlation (DCC) structure of Engle (2002) is employed. The resulting model is referred to as the DCC-MVNCT model, and accounts for (i) volatility clustering, (ii) changing correlations through time, and (iii) the heavy-tailed, skewed, and potentially non-elliptical nature of the returns. It appears that this approach for portfolio construction has not previously been used in the literature, presumably because estimation of the MVNCT was hampered by the slow and numerically challenging evaluation of the density. In the univariate case, the density could be replaced by its closed form expression for the saddlepoint equation without any practical loss in estimation accuracy; see Broda and Paolella (2007) and Paolella (2010). In the multivariate case, however, a saddlepoint approximation has yet to be developed, and unless it also exhibits a closed form saddlepoint, it will require the solution of K nonlinear equations in K unknowns, where K is the number of assets under consideration, and thus will not be useful for any practical value of K. We address this issue and propose a new approximation to this multivariate density that is fast to evaluate and easy to implement. Concerning the overall estimation of the DCC-MVNCT model, a three-step procedure suggests itself, consisting of the usual two steps associated with Engle's DCC model and of an estimate of the (remaining) parameters of the MVNCT from the filtered DCC residuals. Related methods, which likewise invoke properties of quasi maximum likelihood estimation, have been used in the non-Gaussian case, e.g., by Aas et al. (2006), Ku (2008), Jondeau (2010), and Bonato (2012). For computing the MESP the highly useful result of Rockafellar and Uryasev (2000) is applied which avoids the explicit calculation of the expected shortfall during the portfolio optimization. The procedure, however, includes minimizing a numerically evaluated integral expression, such that speed and accuracy are limited. Using the main result of Broda and Paolella (2010), a closed form expression for this integral is derived that enables the fast and accurate computation of large scale minimum ES portfolios using the MVNCT. The approach is general (though it may not always result in a closed form expression) and carries over to all distributions applicable for use with Broda and Paolella (2010), i.e., that permit a stochastic representation in terms of underlying random variables with a tractable moment generating function.

The outline of this paper is as follows. Section 3.2 reviews the idea of Broda and Paolella (2010) and restates the results required in the subsequent derivation of the closed form expression for the MVNCT. Section 3.3 introduces the DCC-MVNCT model, details computational aspects and establishes the closed form expression for use with Rockafellar and Uryasev's result. Section 3.3.1 outlines the empirical results obtained from an extensive out-of-sample forecast study, and compares different performance and quality measures for the DCC and the DCC-MVNCT model. Section 3.4 concludes.

3.2 Saddlepoint approximation of expected shortfall for transformed means

Let X be a random variable with density $f_X(x)$ and cumulant generating function $\mathbb{K}_X(t)$, with $\mathbb{K}_X(t)$ converging in a nonvanishing interval containing the origin. For a given confidence level $q \in (0, 1)$, the expected shortfall for X is defined as

$$\mathrm{ES}^{(q)}(X) \equiv -\mathbb{E}\left[X | X \leq x_q\right] = -\frac{1}{q} \int_{-\infty}^{x_q} x f_X(x) \mathrm{d}x,$$

where x_q refers to the $100q\%$ quantile of X, the Value-at-Risk (VaR), $\mathrm{VaR}^{(q)}(X)$. Alternatively the ES of X can be written as

$$\mathrm{ES}^{(q)}(X) = -\frac{I(x_q)}{F(x_q)}, \quad I(c) \equiv \int_{-\infty}^{c} x f_n(x)\mathrm{d}x,$$

which leads to the idea in Broda and Paolella (2010), to write $I(c)$ as

$$I(c) = cF(c) - \tilde{I}(c), \quad \tilde{I}(c) = \int_{-\infty}^{c} (c-x) f_X(x)\mathrm{d}x, \tag{3.1}$$

and to approximate $F(c)$ and $\tilde{I}(c)$ separately. Different saddlepoint approximations for $F(c)$ and $\tilde{I}(c)$ as well as for expected shortfall,

$$\mathrm{ES}^{(q)}(X) = -\hat{I}(x_q), \quad \hat{I}(c) \equiv \frac{I(c)}{F(c)}, \tag{3.2}$$

are outlined and discussed in Broda and Paolella (2010, Section 2 and 3).

The idea is then generalized in Broda and Paolella (2010, Section 4) to random variables that are functions of two random variables where each of the underlying random variable possesses a tractable moment generating function. This includes random variables with an intractable moment generating function but an appropriate stochastic representation, an example of which is the noncentral t distribution. Using Temme (1982) and Daniels and Young (1991), saddlepoint expressions for transformed (means of) bivariate random vectors are obtained, and the main result, a saddlepoint approximation for expected shortfall of transformed means, is worked out. Theorem 1 restates the main result, where $\mathbb{E}\left[Y_1 \middle| Y_1 < c\right]$ refers to $\hat{I}(c)$.

THEOREM 1. *Let $\boldsymbol{X} = (X_1, X_2)'$ be a bivariate random vector possessing a density and joint cumulant generating function $\mathbb{K}_{\boldsymbol{X}}(\boldsymbol{t})$. Let g be a smooth bijection such that $\boldsymbol{X} = g(\boldsymbol{Y}) = (g_1(\boldsymbol{Y}), g_2(\boldsymbol{Y}))'$, with inverse $\boldsymbol{Y} = (Y_1, Y_2)' = h(\boldsymbol{X}) = (h_1(\boldsymbol{X}), h_2(\boldsymbol{X}))'$. Then*

$$\mathbb{E}\left[Y_1 \mid Y_1 < c\right] \sim (\alpha_0 + \alpha_1) + \frac{\phi(\tilde{w}_c)}{\hat{F}^1(c)} \left((c-\alpha_0)\left(\frac{1}{\tilde{w}_c^3} - \frac{1}{\tilde{u}_c}\right) - \frac{1}{\left(\tilde{\boldsymbol{t}}_c{}' \nabla_{y_1} g(\tilde{\boldsymbol{y}}_c)\right)\tilde{u}_c} + \frac{\alpha_1}{\tilde{u}_c} \right), \tag{3.3}$$

where $\tilde{w}_c = \mathrm{sgn}(c-\alpha_0)\sqrt{2\left(\tilde{\boldsymbol{t}}_c{}' g(\tilde{\boldsymbol{y}}_c) - \mathbb{K}_{\boldsymbol{X}}(\tilde{\boldsymbol{t}}_c)\right)}$, $\tilde{u}_c = \left(\tilde{\boldsymbol{t}}_c' \nabla_{y_1} g(\tilde{\boldsymbol{y}}_c)\right)\tilde{K}_c^{1/2}$, $\tilde{\boldsymbol{y}}_c = (c, \tilde{y}_{2,c})'$, $\alpha_0 = h_1(\mathbb{K}_{\boldsymbol{X}}'(\boldsymbol{0}))$, $\alpha_1 = h^{ij}\kappa_{ij}/2$, h^{ij} and κ_{ij} are the elements of the Hessian of $h_1(\boldsymbol{X})$ at $\mathbb{K}_{\boldsymbol{X}}'(\boldsymbol{0})$ and $\mathbb{K}_{\boldsymbol{X}}''(\boldsymbol{0})$, respectively,

$$\tilde{K}_c = \frac{\det\left(\mathbb{K}_{\boldsymbol{X}}''(\tilde{\boldsymbol{t}}_c)\right)}{\det^2\left(J_g(\tilde{\boldsymbol{y}}_c)\right)} \left[\nabla_{y_2} g(\tilde{\boldsymbol{y}}_c)' \left(\mathbb{K}_{\boldsymbol{X}}''(\tilde{\boldsymbol{t}}_c)\right)^{-1} \nabla_{y_2} g(\tilde{\boldsymbol{y}}_c) + \tilde{\boldsymbol{t}}_c' \nabla_{y_2}^2 g(\tilde{\boldsymbol{y}}_c) \right],$$

$$\hat{F}^1(c) = \Phi(\tilde{w}_c) + \phi(\tilde{w}_c)\left[\tilde{w}_c^{-1} - \tilde{u}_c^{-1}\right], \tag{3.4}$$

and, for each value of c, $\tilde{\boldsymbol{t}}_c$ and $\tilde{y}_{2,c}$ solve the system

$$\mathbb{K}_{\boldsymbol{X}}'(\tilde{\boldsymbol{t}}_c) = g(\tilde{\boldsymbol{y}}_c),$$
$$\tilde{\boldsymbol{t}}_c' \nabla_{y_2} g(\tilde{\boldsymbol{y}}_c) = 0.$$

In order to apply Theorem 1 to the univariate (standardized) singly noncentral t, let $X_1 \sim \mathrm{N}(\gamma, 1)$, and let $X_2 \sim \chi^2(\nu)$ independently of X_2, where γ is the noncentrality parameter, and ν denotes the degrees of freedom. Using $\boldsymbol{X} = g(\boldsymbol{Y}) = (Y_1 Y_2, Y_2^2 \nu)'$, it follows $(Y_1, Y_2)' = g^{-1}(X_1, X_2) = \left(X_1/\sqrt{X_2/\nu}, \sqrt{X_2/\nu}\right)'$, such that Y_1 has a singly noncentral t distribution. The expression obtained for the noncentral t is explicit, straightforward to implement and extremely fast to evaluate.[2]

3.3 Portfolio optimization using the noncentral t

For modeling the distribution of portfolio returns we consider the singly noncentral MVNCT. One of the features of this distribution is that linear combinations of jointly MVNCT distributed random variables remain in the same family of distributions. That is, the distribution of portfolio returns is univariate noncentral t, and allows the use of Theorem 1. We restrict our attention to the singly noncentral case, which is more readily generalized to the multivariate setting. In Broda and Paolella (2007), it was found that the singly noncentral t is sufficiently flexible as a model for the distribution of assets returns, so this assumption is unproblematic.

The stochastic representation of the (singly noncentral) K-variate MVNCT is as follows. Let the K-vector $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{\gamma}, \boldsymbol{\Sigma})$, and the scalar $Y \sim \chi^2(\nu)$, independently of $\boldsymbol{Z}$. Then

$$\boldsymbol{X} = \frac{\boldsymbol{Z}}{\sqrt{Y/\nu}} \sim \mathrm{MVNCT}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are K-vectors of location and noncentrality coefficients, respectively, ν denotes the degrees of freedom, and $\boldsymbol{\Sigma}$ is the dispersion matrix. This distribution has first been considered in Kshirsagar (1961); see also Kotz and Nadarajah (2004, Section 5.1). Its first and second moments are

$$\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu} + \zeta\boldsymbol{\gamma} \quad \text{and} \quad \mathbb{V}[\boldsymbol{X}] = \frac{\nu}{\nu-2}\left(\boldsymbol{\Sigma} + \boldsymbol{\gamma}\boldsymbol{\gamma}'\right) - \zeta^2\boldsymbol{\gamma}\boldsymbol{\gamma}',$$

where $\zeta = \sqrt{\nu/2}\,\Gamma\left((\nu-1)/2\right)/\Gamma(\nu/2)$, see, e.g., Jondeau (2010).

We use the MVNCT in the following portfolio model. Let $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{k,t})'$, $t = 1, \ldots, T$, denote the set of asset returns corresponding to K assets at time t. We assume that

$$\boldsymbol{r}_t = \boldsymbol{m} + \boldsymbol{H}_t^{1/2}\boldsymbol{X}_t, \quad \text{where } \boldsymbol{X}_t \overset{\text{iid}}{\sim} \mathrm{MVNCT}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}), \tag{3.5}$$

and $\boldsymbol{H}_t$ is a time-varying covariance matrix, see below. For computing the matrix square root, we refer to the eigenvalue decomposition $\boldsymbol{H}_t = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}'$ such that $\boldsymbol{H}_t^{1/2} \equiv \boldsymbol{V}\boldsymbol{E}\boldsymbol{V}'$ is symmetric, where $\boldsymbol{E}$ is the diagonal matrix with $\boldsymbol{e}_{ii} = \sqrt{d_{ii}}$. Clearly not all parameters in (3.5) are identified. In order to ensure

[2] More detailed results and information on computational aspects are available from the authors upon request.

identification, we impose analogously to Jondeau (2010) parameter restictions to guarantee that $\boldsymbol{X}_t$ has conditional mean zero and unit covariance[3] by setting

$$\boldsymbol{\mu} = -\zeta\boldsymbol{\gamma} \quad \text{and} \quad \boldsymbol{\Sigma} = (\boldsymbol{I}_{K\times K} + \boldsymbol{\mu}\boldsymbol{\mu}')\frac{\nu-2}{\nu} - \boldsymbol{\gamma}\boldsymbol{\gamma}'.$$

This ensures that $\boldsymbol{H}_t$ is the covariance matrix of $\boldsymbol{r}_t$, provided that $\nu > 2$, which we assume throughout. A further restriction is required to ensure positive definiteness of $\boldsymbol{\Sigma}$. Observe that

$$\boldsymbol{\Sigma} = \frac{\nu-2}{\nu}\boldsymbol{I}_{K\times K} - \left(1 - \frac{\nu-2}{\nu}\zeta^2\right)\boldsymbol{\gamma}\boldsymbol{\gamma}',$$

where $\boldsymbol{\gamma}\boldsymbol{\gamma}'$ is a rank-1 matrix with a single non-zero eigenvalue $\boldsymbol{\gamma}'\boldsymbol{\gamma}$. Hence the positive definiteness of $\boldsymbol{\Sigma}$ is ensured by requiring

$$\boldsymbol{\gamma}'\boldsymbol{\gamma} < \frac{\nu-2}{(1-\zeta^2)\,\nu + 2\zeta^2}.$$

Using these restrictions, the distribution of the vector of asset returns in (3.5) follows as

$$\boldsymbol{r}_t \sim \text{MVNCT}(\tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\gamma}}_t, \nu, \tilde{\boldsymbol{\Sigma}}_t), \tag{3.6}$$

$$\tilde{\boldsymbol{\Sigma}}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}\boldsymbol{H}_t^{1/2}, \quad \tilde{\boldsymbol{\gamma}}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\gamma}, \quad \tilde{\boldsymbol{\mu}}_t = \boldsymbol{m} - \zeta\tilde{\boldsymbol{\gamma}}_t.$$

The (univariate) distribution of the return of a portfolio with weights $\boldsymbol{w}$ implied by (3.6) is

$$R_t = \sum_{i=1}^{K} w_i r_{i,t} \sim \text{MVNCT}\left(\boldsymbol{w}'\tilde{\boldsymbol{\mu}}_t, \boldsymbol{w}'\tilde{\boldsymbol{\gamma}}_t, \nu, \boldsymbol{w}'\tilde{\boldsymbol{\Sigma}}_t\boldsymbol{w}\right), \tag{3.7}$$

where we assume $\boldsymbol{w}'\boldsymbol{1} = 1$ and $0 \leq w_i \leq 1$ (no short selling).[4]

The minimum expected shortfall portfolio (hereafter MESP) problem is now given by

$$\min_{\boldsymbol{w}} \text{ES}^{(q)}(R_t),$$

where q is the VaR level. Then, with the auxiliary function

$$F_q(x, \boldsymbol{w}) = x - \frac{1}{q}\int_{-\infty}^{-x} (r + x) f_{R_t}(\boldsymbol{w}; \boldsymbol{r}_{t+1})\mathrm{d}r, \tag{3.8}$$

where $f_{R_{t+1}}(\boldsymbol{w}; \boldsymbol{r}_{t+1})$ is the predictive density of the portfolio return, Rockafellar and Uryasev (2000) show that

$$\text{ES}^{(q)}(R_t) = \min_x F_q(x, \boldsymbol{w}) \quad \text{and} \quad \text{VaR}^{(q)}(R_t) = \arg\min_x F_q(x, \boldsymbol{w}).$$

[3]In Jondeau (2010, Section 3.2) there appears to be a typographical error in Equation (19) where $\mathbb{E}[\boldsymbol{\varepsilon}_t]$ needs to be replaced by $\mathbb{E}[\boldsymbol{\varepsilon}_t - \boldsymbol{m}]$.

[4]This is different from the univariate parametrization of the previous section. In terms of the latter, $(R_t - \boldsymbol{w}'\boldsymbol{m} - \boldsymbol{w}'\boldsymbol{\mu})/\sigma \sim \text{NCT}(\boldsymbol{w}'\boldsymbol{\gamma}/\sigma, \nu)$, where $\sigma^2 = \boldsymbol{w}'\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}\boldsymbol{H}_t^{1/2}\boldsymbol{w}$.

Accordingly, the MESP problem can be solved by minimizing $F_q(x, \boldsymbol{w})$ jointly over x and $\boldsymbol{w}$. This avoids the explicit calculation of the expected shortfall during the optimization. The resulting procedure delivers the ES, the VaR and the optimal portfolio weights from a single optimization.

For the predicted ES, let $\tilde{\mu}_{t+1} = \boldsymbol{w}'\tilde{\boldsymbol{\mu}}_{t+1}$, $\tilde{\gamma}_{t+1} = \boldsymbol{w}'\tilde{\boldsymbol{\gamma}}_{t+1}$, $\tilde{\nu}_{t+1} = \nu$, and $\tilde{\sigma}_{t+1} = (\boldsymbol{w}'\tilde{\boldsymbol{\Sigma}}_{t+1}\boldsymbol{w})^{1/2}$ denote the parameters of the predictive distribution of the portfolio return R_{t+1}. Letting $c = -x$, Equation (3.8) can be rearranged as

$$F_q(c, \boldsymbol{w}) = \frac{\tilde{I}_n(c)}{q} - c, \text{ where } \tilde{I}_n(c) = \int_{-\infty}^{c} (c - r) f_{R_{t+1}}(\boldsymbol{w}; \boldsymbol{r}_{t+1}) \mathrm{d}r \tag{3.9}$$

is computed based on (3.3), $n = 1$ applies, and

$$f_{R_{t+1}}(\boldsymbol{w}; \boldsymbol{r}_{t+1}) \equiv f_{\mathrm{MVNCT}}((\boldsymbol{w}'\boldsymbol{r}_{t+1} - \tilde{\mu}_{t+1})/\tilde{\sigma}_{t+1}, 0, \tilde{\gamma}_{t+1}/\tilde{\sigma}_{t+1}, \nu, 1)/\tilde{\sigma}_{t+1}$$

denotes the univariate location zero scale one noncentral t density suitable for Theorem 1. The required expression for $\tilde{I}_n(c)$ is given by $c\hat{F}^1(c) - \hat{E}(c)$, where $\hat{E}(c)$ refers to (3.3), and $\hat{F}^1(c)$ to (3.4).[5] Note that $\hat{F}^1(c) = q$ holds iff $c = \mathrm{VaR}^{(q)}(R_{t+1})$ which will not always be true when (3.8) is minimized. By virtue of saddlepoint techniques, the explicit approximation in (3.9) is extremely fast to compute. Figure 3.1 compares the computation times of (3.8) based on numerical integration and (3.8) based on (3.9). As to be expected, the use of the new expression significantly reduces computation times. A minimum improvement of 21% and a maximum improvement of 80% is achieved for portfolios up to 500 assets; for 30 assets the factor is approximately 4. For larger portfolios the increase in speed seems to become limited by other factors than the execution times for evaluating the ES.

Considering the DCC filter of Engle (2002) for modeling the evolution of $\boldsymbol{H}_t$, the DCC-MVNCT model corresponding to (3.5) emerges. For estimating the model, the two-step estimation procedure of the DCC is augmented by adding a third step that fits the MVNCT to the standardized and decorrelated DCC residuals. The suggested procedure works as follows:

(i) For each of the K univariate time series, fit univariate GARCH models with Gaussian innovations to the residuals $\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{r}_t - \widehat{\boldsymbol{m}}$, such that estimates of the means, $\widehat{m}_i$, and of the time-varying variances, $\widehat{\sigma}^2_{i,t}$, are obtained (estimated jointly).

(ii) Based on the standardized residuals $\widehat{u}_{i,t} = (r_{i,t} - \widehat{m}_i)/\widehat{\sigma}_{i,t}$ compute the Engle's DCC filter and obtain estimates of time-varying covariance structure $\widehat{\mathbf{H}}_t$.

(iii) Estimate the remaining MVNCT parameters, $\boldsymbol{\gamma}$ and ν, using the standardized returns $\widehat{\boldsymbol{H}}_t^{-1/2}(\boldsymbol{r}_t - \widehat{\boldsymbol{m}})$, and respecting the parameter restrictions. Finally, construct the full model as given in (3.6).

[5] To be precise, $\tilde{\sigma}_{t+1}\tilde{I}_n(c)$ is used with the SPA for $f_{\mathrm{MVNCT}}((\boldsymbol{w}'\boldsymbol{r}_{t+1} - \tilde{\mu}_{t+1})/\tilde{\sigma}_{t+1}, 0, \tilde{\gamma}_{t+1}/\tilde{\sigma}_{t+1}, \nu, 1)$.

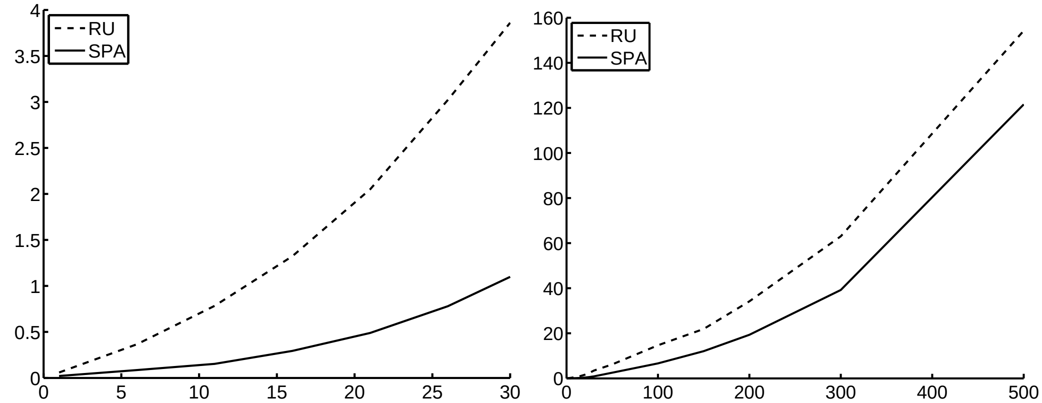


Figure 3.1: Average execution times in seconds for computing the minimum expected shortfall portfolio for different numbers of assets, K. RU refers to (3.8) using Matlab's quadgk routine for numerical integration. SPA refers to (3.8) using the closed form expression in (3.9). For each K, results are based on the same set of randomly generated MVNCT distributions and starting values. For minimizing (3.8) under no short selling constraints, we use Matlab's fmincon routine, i.e., its interior-point method. Both methods yield the same results.

Density forecasts are obtained when the DCC prediction $\widehat{\boldsymbol{H}}_{t+1}$ is plugged into (3.6). Details on the numerical maximization of the likelihood function required in step (iii) are provided in Appendix A.

3.3.1 Empirical results

To demonstrate the method, we consider the daily percentage log-returns of the 30 components of the DJIA from Wharton/CRSP (as used in April 2013) and evaluate the model performances by applying each model to a series of $N = 2500$ moving windows of 1000 observations (roughly 4 years of data). For each window the one-day-ahead density forecast is obtained and optimal portfolio weights corresponding to the MESP are computed. The out-of-sample forecast period ranges from January 2003 to December 2012, covering 10 years of data. The models under study are the Gaussian DCC and the DCC-MVNCT model, as well as the central case ($\gamma = 0$) of the latter, called DCC-MVT. Model parameters and the portfolio weights are updated on a daily basis.

3.3.1.1 Methodology

We consider five different perspectives for out-of-sample forecast comparison.

First, we compare portfolios with regard to the achieved returns and different risk-adjusted performance measures. Working with the N realized portfolio returns, $\tilde{r}_{t+1} = \boldsymbol{w}_{t+1}^{\star}{}'\boldsymbol{r}_{t+1}$, $\tilde{\boldsymbol{r}} = (\tilde{r}_1, \ldots, \tilde{r}_N)'$, where $\boldsymbol{w}_{t+1}^{\star}$ refers either to the (predictive) min-ES or the equally-weighted portfolio based on the previous 1000 returns ending with date t, we report the annualized return, standard deviation and Sharpe ratio

(Sharpe, 1966), the Sortino ratio (Sortino and Price, 1994), and, for a confidence level of $q = 1\%$, the STARR ratio (Stable Tail Adjusted Return Ratio; Martin et al., 2003), as well as what we term the realized average VaR exceedance (RAVE) and the conditional realized average VaR exceedance (CRAVE). The latter two measures are defined as

$$\text{RAVE}_q(\tilde{\boldsymbol{r}}) = \frac{N^{-1}\sum_{t=1}^{N} \tilde{r}_t \mathbb{1}\left\{\tilde{r}_t < Q_q^*\left(\tilde{\boldsymbol{r}}\right)\right\}}{N^{-1}\sum_{t=1}^{N} \mathbb{1}\left\{\tilde{r}_t < Q_q^*\left(\tilde{\boldsymbol{r}}\right)\right\}} \quad \text{and } \text{CRAVE}_q(\tilde{\boldsymbol{r}} \mid \tilde{\boldsymbol{z}}) = \frac{N^{-1}\sum_{t=1}^{N} \tilde{r}_t \mathbb{1}\left\{\tilde{z}_t < Q_q^*\left(\tilde{\boldsymbol{z}}\right)\right\}}{N^{-1}\sum_{t=1}^{N} \mathbb{1}\left\{\tilde{z}_t < Q_q^*\left(\tilde{\boldsymbol{z}}\right)\right\}},$$

where $\mathbb{1}$ is the indicator function, Q_q^* is the empirical quantile function, and $\tilde{z}_t = \left(\tilde{r}_t - \widehat{\mu}\right)/\widehat{\sigma}_t$ with $\widehat{\mu}$ and $\widehat{\sigma}_t$ being estimated by fitting an auxiliary model to $\tilde{\boldsymbol{r}}$.[6] The RAVE can be seen as an unconditional empirical measure of expected shortfall, while the CRAVE constitutes a conditional version by introducing an auxiliary model that accounts for the dynamic nature of financial returns. In general, the essence of CRAVE is to lead to a more reasonable measurement of risk by the embedding of market dynamics. This, for example, is accomplished if the auxiliary model allows for time-varying volatility such that large returns in times of high (low) market volatility become less (more) likely to appear as VaR violations, compared to the RAVE. For reasons of comparison, we require the auxiliary model to be independent of the portfolio model, and show results for different GARCH$(1,1)$ models, i.e., the t-GARCH and the NCT-GARCH model based on the central t and the noncentral t distribution, respectively, as well as the MixNormal(3)-GARCH model introduced in Haas et al. (2004) and Alexander and Lazar (2006). Details on the estimation of the latter are given in Broda (2013).

Second, we compare forecast qualities based on likelihood values. Let $\hat{q}_{t+1} = \log \hat{f}_{t+1|\mathcal{F}_t}(\tilde{r}_{t+1}; \hat{\boldsymbol{\theta}}_t)$, be the realized predictive log-likelihood value, where $\hat{f}_{t+1|\mathcal{F}_t}$ refers to the predictive density at time $t+1$ conditional on $\mathcal{F}_t$, that represents the information set available at time t. Further let $\hat{\boldsymbol{q}} = (\hat{q}_1, \ldots, \hat{q}_N)$ denote the time-ordered vector of all N forecasts obtained from the out-of-sample forecast exercise. An obvious approach is to compare the sums of $\hat{\boldsymbol{q}}$ among the different models under study, where the best performing model is characterized by the highest value. A test for comparing forecasts based on $\hat{\boldsymbol{q}}$ is given in Diebold and Mariano (1995). Observe that each model results in two forecasts, namely the one from the multivariate model for the K-vector of all asset returns and the one from the (derived) univariate model for the portfolio return conditional on $\boldsymbol{w}_{t+1}^{\star}$. While the full multivariate distributional forecast is of general interest, the distribution of the linear combination of the marginals is of particular one as it assembles the portfolio. We report test results for both.

Third, we compare forecast qualities based on cumulative distribution function (cdf) values, referring to the probability integral transform (see, e.g., Neyman, 1937; and Rosenblatt, 1952). Analo-

[6] Another approach is to construct an improved RAVE measure by using the VaR estimates of the auxiliary model instead of employing the empirical quantile of the standardized data. This, however, leads to quite similar results.

gously to the last paragraph, we consider the realized predictive cdf value, $\hat{p}_{t+1} = \log \hat{F}_{t+1|\mathcal{F}_t}(\tilde{r}_{t+1})$, $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_N)$. If the true model is applied, it is well-known that the $\hat{p}_i$ are iid uniformly distributed, which can be tested. We briefly remark that there is an ongoing discussion, or uncertainty, which approach to performance evaluation is superior, (a) using realized predictive likelihood values, or (b) using realized predictive cdf values; see, e.g., Geweke and Amisano (2009). As detailed in Broda et al. (2013, Section 4.3), we test for uniformity of the $\hat{p}_i$ using the Anderson-Darling, the Cramér-von Mises, the Kolmogorov-Smirnov and the Pearson test, where the null hypothesis assumes a uniform distribution on the unit interval. Likewise results are reported for the Jarque-Bera and the Shapiro-Wilk test testing for normality of the transformed values $\Phi^{-1}(\hat{p}_i)$, where Φ is the cdf of the standard normal distribution. Furthermore we test for serial correlation in $\hat{\boldsymbol{p}}$ and report Ljung-Box test results.

Fourth, we compare conditional coverage qualities using VaR measures dedicated to the left tail. These quantities, also formed from the realized predictive cdf values, are possibly of greater interest from a risk management perspective than the distributional tests in the former paragraph. We show VaR coverage probabilities and the integrated root mean squared error over the left tail. The empirical coverage probability (in percent) is given by $100F_N(q)$, where F_N is the empirical distribution function based on $\hat{\boldsymbol{p}}$, while the integrated error refers to a truncated Cramér-von Mises test statistic and gives the coverage error over the lower VaR levels up to the $100q\%$ level,

$$\sqrt{\frac{1}{h}\sum_{m=1}^{h}\left(100\frac{2m-1}{2N} - 100\hat{p}_m^{[s]}\right)^2},$$

where $h = \lceil qN \rceil$, and $\hat{\boldsymbol{p}}^{[s]}$ refers to $\hat{\boldsymbol{p}}$ sorted in increasing order. Additional details on the latter approach are given in Kuester et al. (2006) and Broda et al. (2013).

Fifth, we compare conditional coverage qualities based on the time-ordered sequence of realized predictive VaR violations, $\boldsymbol{v} = (v_1, \ldots, v_N)$, $v_i = \mathbb{1}\{\tilde{r}_i \leq \mathrm{VaR}^{(q)}(R_i)\}$. It is well-known that the v_i are iid Bernoulli(q) under the correct model. We report the Christoffersen (1998) test results for the overall test of conditional coverage as well as the individual tests for unconditional coverage and independence, i.e., $\mathrm{LR}_{\mathrm{CC}} = \mathrm{LR}_{\mathrm{UC}} + \mathrm{LR}_{\mathrm{IND}}$.

3.3.2 Discussion of results

Briefly, the out-of-sample results obtained for the Gaussian DCC, the DCC-MVT and the DCC-MVNCT model are very close in portfolio performance. In particular, the performances of the latter two models can hardly be distinguished from each other, whereas the Gaussian DCC model slightly underperforms. In terms of forecast quality, the results are more distinct, but likewise with the non-Gaussian models ahead of the Gaussian one.

At first, we look at the risk and return performance of the min-ES portfolios (without short selling) obtained from the different models. For comparison, we also consider the equally-weighted portfolio. The annualized performance measures, discussed in the previous section, for the out-of-sample period 2003–2012 are shown in Table 3.2. Annual performances are presented in Figure 3.3. By means of chance, and always exhibiting the highest annual variance, the equally-weighted portfolio outperforms the min-ES portfolios slightly (by less than 1%) at the 10-year annualized return. All three DCC models yield about the same annualized return with differences being marginal. Even on the daily frequency the min-ES portfolios perform the same, as can be seen from the cumulative returns and differences thereof in Figure 3.4. As this is an unexpected result, we take a closer look at the underlying models and compare quantities (mean, variance and median, as well as the 1% VaR and ES) that characterize the density predictions, based on which the portfolios are computed. For the Gaussian DCC and the DCC-MVNCT model, Figure 3.4 and 3.5 show these quantities over time. As it turns out, mean and variance predictions are virtually identical whereas the predictions of median, 1% VaR and 1% ES differ significantly; being higher for the DCC-MVNCT model. From this, the density predictions of the DCC-MVNCT model can be expected to have heavy tails (heavier than normal) and to be of asymmetric shape. This is confirmed in Figure 3.6 and 3.7, respectively, from the obtained estimates of the degrees of freedom and of the noncentrality parameters. The estimated degrees of freedom are found between 8 and 12, and the estimated noncentrality coefficients between -0.1 and 0.7. From Figure 3.7 it also becomes evident that the non-elliptical tail behavior varies over time and is amplified in times of financial crisis. Turning back to the mean and variance predictions, a similar picture presents itself for the equally-weighted portfolio, as shown in Figure 3.8 and 3.9. Likewise here, the predictions of mean and variance remain unaffected when the distributional assumption is changed. We attribute this to the use of the three step estimation procedure that prevents much of the interaction between the various shape parameters of the DCC-MVNCT model, see below. An alternative estimation approach however is pending.

For the risk-adjusted performance measures presented in Table 3.2, the min-ES portfolios perform almost indistinguishably, and differences are negligibly. Nevertheless, the realized returns of the min-ES portfolio derived from the DCC-MVNCT predictions are slightly favored across the more sophisticated risk-adjusted measures, e.g., Sharpe ratio and CRAVE. The equally-weighted portfolio instead is rated with the highest risk by all measures.

Similarly, the DCC-MVNCT is slightly favored in terms of realized predictive likelihood values, see Table 3.2. According to the Diebold-Mariano test, however, the discrepancies are insignificant. The Gaussian DCC model yields the worst results.

With the DCC-MVT model slightly ahead of the DCC-MVNCT, this picture carries over to the forecast quality measures reported in Table 3.3 and 3.4. Here the prediction quality is shown on the univariate portfolio level, i.e., by evaluating the forecast quality for the univariate density predictions (derived from the multivariate density forecasts conditional on the portfolio weights). For all models under study, forecast qualities are evaluated with respect to the min-ES portfolios as well as to the equally-weighted portfolio. Table 3.3 gives the results for the different min-ES portfolios, while the results in Table 3.4 refer to the equally-weighted one. In both cases, the majority of results is in favor of the DCC-MVT model, closely followed by the DCC-MVNCT. The Gaussian DCC model performs the worst. Judging from the similar results of the DCC-MVT and the DCC-MVNCT model, we conjecture that the full potential of the DCC-MVNCT model lies idle by virtue of the impeding effects of the three step estimation procedure on the parameters, e.g., by separating the estimation of variance and noncentrality related parameters. We conjecture further that the additional K parameters for modelling non-ellipticity may as well introduce some form of overfitting compared to the DCC-MVT model. Hence, one may also want to consider to reduce the number of these parameters or, alternatively, to use a shrinkage approach to shrink them towards zero (the central case) in non-crisis times. This, however, goes beyond the scope of this paper.

Putting all results together, the DCC-MVNCT model, estimated via the three step procedure outlined in Section 3.3 and conditional on the used data, can be said to outperform the classic DCC model in forecasting quality and portfolio performance, if for the latter risk-adjusted performance measures are considered. The results also indicate that it performs better than its elliptical case, the DCC-MVT model, even though only very slightly, in portfolio performance as well as, on the multivariate level, density forecast quality.

3.4 Conclusion

Considering the multivariate singly noncentral t distribution, the present paper gives an application in portfolio optimization building upon the work on saddlepoint approximation for expected shortfall of transformed means in Broda and Paolella (2010). A connection of the main result devised therein to the optimization of minimum expected shortfall portfolios using Rockafellar and Uryasev's result is outlined and is found to significantly reduce computation times in the optimization of large scale portfolios. For modeling the distribution of all asset returns a multivariate GARCH model is proposed that considers the noncentral Student's t distribution in combination with the DCC filter of Engle (2002). The resulting DCC-MVNCT model captures most stylized facts of asset returns, including time varying volatility and correlation, fat tails and asymmetry, as well as non-ellipticity. Concerning its estimation, a simple three step procedure is suggested that makes use of the newly developed approximation to the multivariate noncentral t density, that otherwise is not available. Model performance and prediction quality are assessed by an out-of-sample forecast study based on the components of the DJIA30 index. The DCC-MVNCT model is found to improve the forecasting quality compared to the Gaussian DCC model. Compared to its central (elliptical) case, however, the potential advantage of the DCC-MVNCT model of being able to capture non-ellipticity remains unclear, and noncentralicy is only seen to increase during times of crisis. This may justify a shrinkage approach, to be considered in future research, in which the DCC-MVNCT model is shrunk towards its central case when markets are calm. Another point that can be challenged regarding the empirical results is the three step estimation procedure that limits the free interplay of parameters, in particular of those being responsible for modeling non-normality. It stands to reason that another estimation approach will lead to different estimates and empirical results, which are to be assumed to be more in favor of the DCC-MVNCT model. A most promising approach in that direction, that is applicable to the MVNCT distribution, is found in Paolella and Polak (2013), and constitutes future work.

Appendix

A Approximating the MVNCT

We consider Kshirsagar's multivariate NCT (MVNCT) distribution; e.g., see Kotz and Nadarajah (2004, Section 5.1). Let $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{\gamma}, \boldsymbol{\Sigma})$ and $X \sim \chi^2(\nu)$ be independent random variables, then

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{Z}/\sqrt{X/\nu} \sim \mathrm{MVNCT}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}) \tag{3.10}$$

follows a K-variate MVNCT with density function

$$\begin{aligned} f_{\boldsymbol{Y}}\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\gamma},\nu,\boldsymbol{\Sigma}\right) &= \frac{\Gamma\left(\left(\nu+K\right)/2\right)}{\left(\pi\nu\right)^{K/2}\Gamma\left(\nu/2\right)\sqrt{\det\left(\boldsymbol{\Sigma}\right)}} \exp\left\{-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\right\}\left(\frac{\nu}{\nu+\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}}\right)^{(\nu+K)/2} \\ &\quad \times \sum_{k=0}^{\infty}\frac{2^{k/2}\Gamma\left(\left(\nu+K+k\right)/2\right)}{k!\Gamma\left(\left(\nu+K\right)/2\right)}\left(\frac{\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}{\sqrt{\nu+\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}}}\right)^{k}, \end{aligned}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ is the vector of location coefficients, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ is the vector of noncentrality parameters, $\boldsymbol{\Sigma} = (\sigma_i\sigma_j)_{1\le i,j\le K}$ denotes the covariance matrix of $\boldsymbol{Z}$ (hereafter also called the dispersion matrix of $\boldsymbol{Y}$), ν is the degrees of freedom parameter of X, and $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{\mu}$. Ignoring the evaluation point $\boldsymbol{x}$, the density possesses $2K + 1 + K(K+1)/2$ parameters.

For estimating the MVNCT the standard maximum likelihood estimator (MLE) could be employed. However, bearing in mind the fast (quadratically) growing number of parameters relative to K, the curse of dimensionality restricts ML estimations to small K when all model parameter are to be estimated. For $K = 3$ the model comprises 13 parameters, while the number doubles to 26 for $K = 5$. Considering the DJIA30 components, standard ML estimations of the 30-dimensional MVNCT (526 parameters) turn out to be computationally infeasible—unless the number of parameters can somehow be restricted, e.g., by imposing parameter constraints.

A.1 Direct Approximation

The evaluation of Kshirsagar's K-variate NCT density, e.g., see Kshirsagar (1961) and Kotz and Nadarajah (2004, Section 5.1), is a non-trivial problem as closed form expressions are not available. We suggest a fast, reliable and accurate (accurate for the desired range of applications) approximation procedure that avoids the common (numerical) problems associated with the infinite sum representation of the noncentrality part of the density.

Let $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{\gamma}, \boldsymbol{\Sigma})$ and $X \sim \chi^2(\nu)$ be independent random variables, then

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{Z}/\sqrt{X/\nu} \sim \mathrm{MVNCT}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}) \tag{3.11}$$

follows Kshirsagar's K-variate NCT distribution with density function

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}) &= \frac{\Gamma\left((\nu+K)/2\right)}{(\pi\nu)^{K/2}\,\Gamma(\nu/2)\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\right\}\left(\frac{\nu}{\nu+\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}}\right)^{(\nu+K)/2} && (3.12)\\ &\quad \times \sum_{k=0}^{\infty} g_k(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}), \\ g_k(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}) &= \frac{2^{k/2}\Gamma\left((\nu+K+k)/2\right)}{k!\,\Gamma\left((\nu+K)/2\right)}\left(\frac{\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}{\sqrt{\nu+\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}}}\right)^{k}, && (3.13)\end{aligned}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ is the vector of location coefficients, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ is the vector of noncentrality parameters, $\boldsymbol{\Sigma} = (\sigma_i\sigma_j)_{1\le i,j\le K}$ denotes the covariance matrix of $\boldsymbol{Z}$, ν is the degrees of freedom parameter of X, and $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{\mu}$.

For stabilizing the computation of the infinite sum, it appears beneficial to work with the log of (3.12),

$$\begin{aligned} \log f_{\boldsymbol{Y}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}) &= \log\Gamma\left(\frac{\nu+K}{2}\right) - \frac{K\log(\pi\nu)}{2} - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{\log\det(\boldsymbol{\Sigma})}{2} && (3.14)\\ &\quad + \frac{\nu+K}{2}\left(\log(\nu) - \log\left(\nu+\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)\right) && (3.15)\\ &\quad - \frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}{2} + \log\sum_{k=0}^{\infty} g_k(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}). && (3.16)\end{aligned}$$

The use of logarithms for linearizing products is well-known to increase both numerical robustness and accuracy, e.g., by preventing numerical under- and overflows. Besides

$$\sum_{k=0}^{\infty} g_k \left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}\right), \tag{3.17}$$

the evaluation of $f_{\boldsymbol{Y}}$ to machine precision is straightforward using standard numerical toolkits. Unfortunately a closed form solution for (3.17) is not available (to the best of our knowledge). It however becomes evident that $f_{\boldsymbol{Y}}$ consists of two different kinds of terms. The terms in (3.14) and (3.15) give rise to the density of the multivariate *central* t (MVT), while those in (3.16) only contribute (by adding noncentrality) if $\boldsymbol{\gamma} \neq \mathbf{0}$. We make use of this property and suggest the following approximation to $f_{\boldsymbol{Y}}$.

Let $\boldsymbol{x}$ be a point on the support of $\boldsymbol{Y}$, and let ϵ, $\epsilon > 0$, be some small threshold value, e.g., the machine precision. The approximation works as follows. First, the MVT density, $f_{\boldsymbol{Y}}^{\gamma=0}$, namely (3.14) and (3.15), is evaluated to machine precision. Then, based on the achieved likelihood value we decide whether to evaluate (3.16) or not, and compute the noncentrality part only if $f_{\boldsymbol{Y}}^{\gamma=0}(\boldsymbol{x}) \geq \epsilon$. This is of particular importance as computations times of (3.16) tend to increase tremendously for (distant) evaluation points that have an almost zero likelihood. Finally, the approximated value, $\hat{f}_{\boldsymbol{Y}}(\boldsymbol{x})$, is returned. By construction this approximation involves an error in the outer tail area where the computation of (3.16) is disregarded. Observe however that $f_{\boldsymbol{Y}}$ will anyway evaluate to a likelihood value close to zero here. Neglecting special cases of extreme noncentrality (which is taken to be a reasonable assumption for financial returns), the approximation comes with no practical loss in accuracy, and the approximation error is small to negligible, depending on ϵ. We suggest machine precision for ϵ. To further tighten the approximation of (3.17), the following results are used.

Let $(g_k)_{k=0,\ldots}$ denote the series of summands. Then, (i) $g_0 = 1$; (ii) g_k is oscillating when $\kappa = \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$ has a negative sign; (iii) $g_k \to 0$ if $k \to \infty$; (iv) series $(h_k)_{k=0,1,\ldots} = (|g_k|)_{k=0,1,\ldots}$ has a global maximum; (v) the infinite sum converges with respect to some reasonable stopping condition within a finite number of summands; (vi) the infinite sum can be accurately approximated without the numerical under- and overflow issues of the naive approach. An illustration of g_k is given in Figure 3.2.

(i) and (ii) are trivial. (iii) Let

$$\nabla_k = 2^{k/2}\Gamma\left(\left(\nu + K + k\right)/2\right) \left(\frac{\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}{\sqrt{\nu + \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}}}\right)^k, \quad \text{and} \quad \Delta_k = k!\Gamma\left(\left(\nu + K\right)/2\right),$$

denote numerator and denominator, respectively, of g_k as functions of k. Clearly the denominator exhibits a higher growth rate than the numerator. That is, Δ_k outweighs ∇_k as k increases, and $g_k \to 0$ in the limit as $k \to \infty$. (iv) Analogously to (ii) we look at $h_k = |g_k|$ and consider the absolute value of ∇_k and Δ_k. As it turns out $|\Delta_k|$ is a monotonically increasing function in k, while $|\nabla_k|$ can be monotonically either decreasing if $|\kappa| < 1$, or is increasing if $|\kappa| \geq 1$. From the monotonicity of $|\Delta_k|$ and $|\nabla_k|$ it

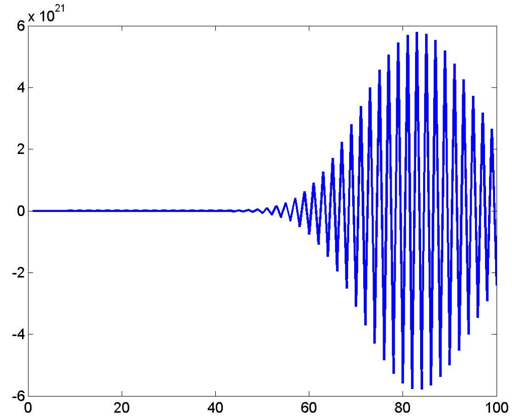


Figure 3.2: The panels show the first 101 terms of g_k (left) and h_k (right), respectively, for exemplary parameters. The left panel depicts an example where g_k is oscillating, while the right panel gives another ones on h_k.

follows that h_k either takes its maximum at $k = 0$ if $|\nabla_k|$ is decreasing, or starts with $h_0 = 1$, takes a maximum at an unknown k and then declines towards zero, if $|\nabla_k|$ is increasing. In both cases, h_k has a global maximum. (v) Convergence requires a stopping condition. Recalling (iii), the summands vanish as k increases. Therefore, the infinite sum can be truncated at $k = k^\star$ with $h_{k^\star} \leq \epsilon$, where $\epsilon > 0$ is an absolute threshold. Alternatively, the sum can be truncated at the first summand that does not significantly contribute to the sum, i.e., at index $k = k^\star$ with $g_{k^\star} / \sum_{j=0}^{k^\star-1} g_j \leq \epsilon$. (iv) Very large values of g_k, e.g., as triggered by large values of ν, can break the numerical limitations when the sum $g_k + \sum_{i=1}^{k-1} g_i$ becomes large. To cope with this problem, the sum is log transformed using the identity

$$\log\left(\exp\{a\} + \sum_i \exp\{b_i\}\right) = a + \log\left(1 + \sum_i \mathfrak{Re}\exp\{b_i - a\}\right),$$

where $a, b_i \in \mathbb{R}$. Now, (3.17) is computed in log scale as

$$s_{k+1} = s_k + \log\left(1 + \mathfrak{Re}\exp\{\log g_k - s_k\}\right), \tag{3.18}$$

where $s_0 = \log g_0$ and

$$\begin{aligned} \log g_k(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \nu, \boldsymbol{\Sigma}) &= \frac{k\log(2)}{2} \log\Gamma\left(\frac{\nu+K+k}{2}\right) - \log\Gamma(k+1) - \log\Gamma\left(\frac{\nu+K}{2}\right) \\ &\quad + k\log\left((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\right) - \frac{k\log\left(\nu + \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)}{2}. \end{aligned}$$

That is, $s_k = \log\sum_{i=0}^{k} g_i$.

Briefly summarized, the benefit of using the “cut-off” approximation based on the central case is a large speed increase in the evaluation of $\hat{f}_{\boldsymbol{Y}}$ as the evaluation of (3.17) can be quite time consuming (if

$\boldsymbol{\gamma} \neq \mathbf{0}$); while the advantage of using (3.18) is a greatly improved numerical robustness and accuracy. The resulting approximation of the multivariate (singly) noncentral t density,

$$\hat{f}_{\boldsymbol{Y}}^{\mathrm{MVNCT}} = f_{\boldsymbol{Y}}^{\mathrm{MVT}} + g_{\boldsymbol{Y}}^{\mathrm{MVNCT}} \mathbb{1}\{f_{\boldsymbol{Y}}^{\mathrm{MVT}} \geq \epsilon\}$$

where $\mathbb{1}$ is the indicator function, $f_{\boldsymbol{Y}}^{\mathrm{MVT}}$ refers to (3.14) and (3.15), and $g_{\boldsymbol{Y}}^{\mathrm{MVNCT}}$ refers to (3.16) computed via (3.18), is robust and quick to evaluate, and therefore, suitable for use in maximum likelihood estimation.

B Figures and Tables

Figures and tables are provided on following pages.

performance measure	min-ES			1/K
	Gaussian DCC	DCC-MVT	DCC-MVNCT	
annualized return	0.0713	0.0711	0.0714	**0.0753**
annualized standard deviation	**0.1459**	0.1459	0.1460	0.2083
annualized Sharpe ratio	0.4886	0.4873	**0.4892**	0.3617
annualized Sortino ratio	0.6908	0.6889	**0.6916**	0.5077
annualized 1% STARR ratio	0.0071	0.0071	**0.0071**	0.0052
annualized 1% RAVE	**10.0110**	10.0195	10.0178	14.5737
annualized 1% CRAVE T-GARCH	6.3043	6.3158	**6.2921**	7.8609
annualized 1% CRAVE NCT-GARCH	6.4141	6.4261	**6.3358**	8.0393
annualized 1% CRAVE MixN(3)-GARCH	6.3761	6.5233	**6.3582**	8.1150

Table 3.1: Annualized performance measures on 10 years (Jan 2003 to Dec 2012) of minimum expected shortfall and equally-weighted portfolio returns. The out-of-sample performance is measured using the realized portfolio returns (non-percentage log-returns) based on the optimal portfolio weights computed from the one-day-ahead density forecasts. Model parameters and portfolio weights (in case of the min-ES portfolio) are updated at every step. For the Gaussian DCC model the usual two-step estimator is used, while for the DCC-MVT and the DCC-MVNCT model the three-step estimation (see Section 3.3) is employed. Entries in boldface denote the best results.

forecast quality measure	Gaussian DCC	DCC-MVT	DCC-MVNCT
average realized predictive log-likelihood for the multivariate density of the K-vector of portfolio returns	-47.310	-45.402	**-45.315**
average realized predictive log-likelihood for the univariate density of the equally-weighted portfolio return	-1.423	**-1.402**	-1.402
average realized predictive log-likelihood for the univariate density of the minimum expected shortfall portfolio return	-1.155**	**-1.115**	-1.116

Table 3.2: Average realized predictive log-likelihood values for 10 years (dating back from Dec 2012) of DJIA30 component returns (as of April 2013). The shown values are obtained from (i) the multivariate predicted density of the vector of portfolio returns and (ii) the (derived) univariate predicted density of the portfolio return (conditional on the vector of portfolio weights). Out-of-sample performance is measured by evaluating the predicted density forecasts at the realized data. Diebold-Mariano test results (H_0: both models forecast equally well) are given in form of ***, **, and *, denoting significance at the 1%, 5%, and 10% level, respectively, relative to the best performing model. Entries in boldface denote the best performer.

forecast quality measure	Gaussian DCC	DCC-MVT	DCC-MVNCT
Sum of realized predictive log-likelihood values	-2886.42**	**-2787.80**	-2789.11
Diebold-Mariano	1.81**	**0.00**	0.03
Anderson-Darling	**6.88*****	8.17***	8.95***
Cramér-von Mises (scaled up by factor 1000)	0.79***	**0.50*****	0.50***
Kolmogorov-Smirnov (scaled up by factor 1000)	**28.42****	31.68**	34.13***
Pearson (50 bins)	125.20***	**113.62*****	119.84***
Jarque-Bera	724.48***	15.24***	**13.82*****
Shapiro-Wilk (transformed by $1000(1-\text{SW})$)	22.28***	2.00***	**1.87*****
Ljung-Box (20 lags)	39.82***	**39.53*****	39.62***
1% VaR coverage percentage	3.24	**2.72**	2.77
1% Integrated root mean squared VaR pred. error	0.55	**0.43**	0.43
1% Unconditional Coverage, LR_{UC}	79.77***	**50.87*****	52.92***
1% Independence, IND_{UC}	0.67	0.64	**0.57**
1% Conditional Coverage, LR_{CC}	80.43***	**51.51*****	53.50***

Table 3.3: Out-of-sample forecast performance results for the **minimum expected shortfall** portfolio, covering 10 years of realized returns data (2003–2012). Performance is measured using the realized predictive (univariate) density and distribution values from evaluating their one-day-ahead forecasts at the realized returns. For the DCC model the usual two-step estimator is applied, while for the DCC-MVT and the DCC-MVNCT model the three-step estimation (see Section 3.3) is used. Entries in boldface denote the best outcome *conditional the optimal min-ES portfolio that corresponds to the particular model*.

forecast quality measure	Gaussian DCC	DCC-MVT	DCC-MVNCT
Sum of realized predictive log-likelihood values	-3558.18	**-3505.64**	-3506.16
Diebold-Mariano	1.01	**0.00**	0.01
Anderson-Darling	9.88***	**3.66****	3.93***
Cramér-von Mises (scaled up by factor 1000)	0.51***	**0.16***	0.17*
Kolmogorov-Smirnov (scaled up by factor 1000)	42.03***	31.25**	**26.45***
Pearson (50 bins)	139.02***	**80.06*****	85.44***
Jarque-Bera	308.21***	26.03***	**25.19*****
Shapiro-Wilk (transformed by $1000(1-\text{SW})$)	17.35***	4.27***	**4.22*****
Ljung-Box (20 lags)	31.59**	**30.78***	30.83*
1% VaR coverage percentage	2.07	**1.66**	1.72
1% Integrated root mean squared VaR pred. error	0.44	**0.23**	0.24
1% Unconditional Coverage, LR_{UC}	22.48***	**8.68*****	10.79***
1% Independence, IND_{UC}	2.21	**1.37**	1.51
1% Conditional Coverage, LR_{CC}	24.69***	**10.05*****	12.29***

Table 3.4: Out-of-sample forecast performance results for the **equally-weighted** portfolio, covering 10 years of realized returns data (2003–2012). Performance is measured using the realized predictive (univariate) density and distribution values from evaluating their one-day-ahead forecasts at the realized returns. For the DCC model the usual two-step estimator is employed, while for the DCC-MVT and the DCC-MVNCT model the three-step estimation (see Section 3.3) is taken. Entries in boldface denote the best results.

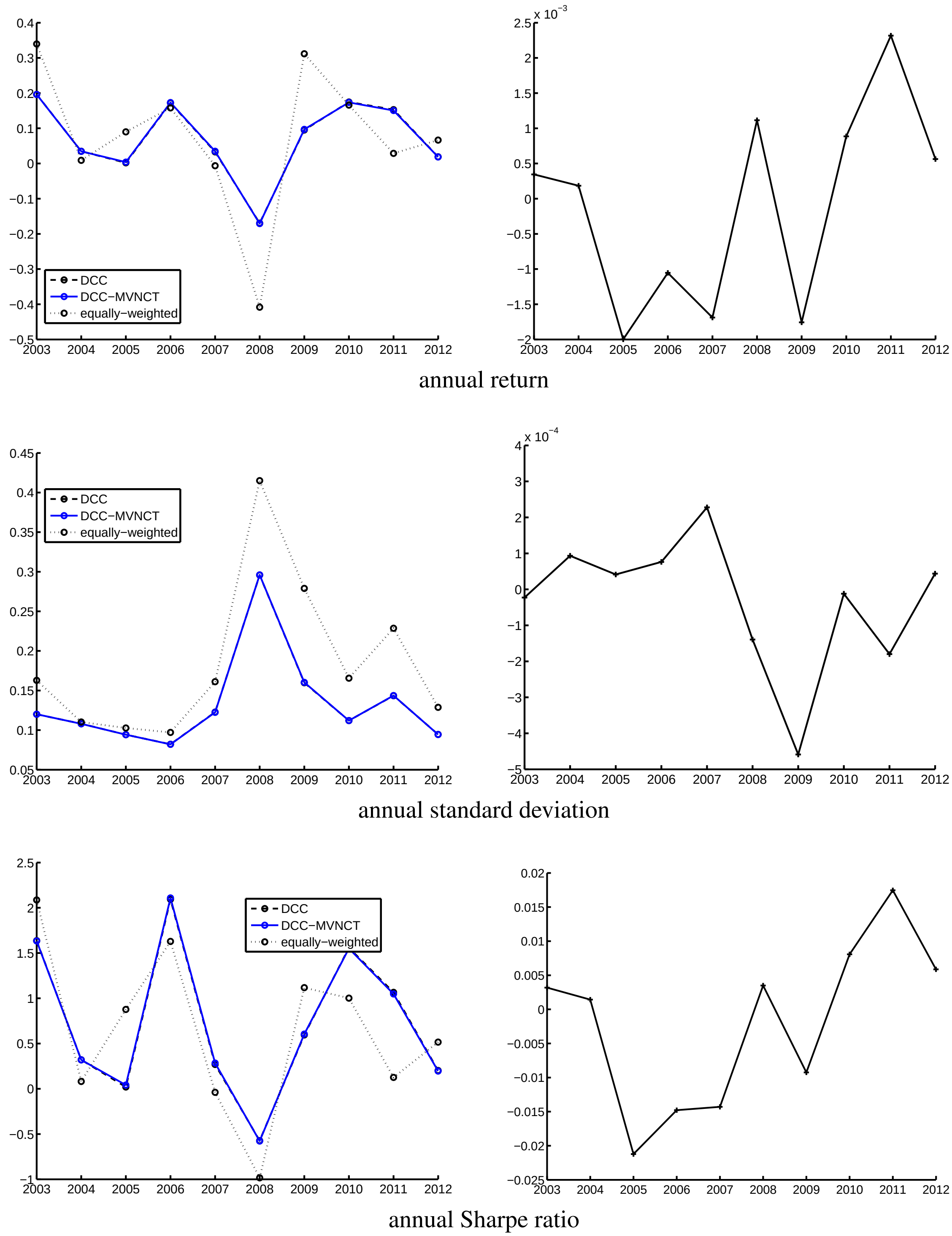


Figure 3.3: Performance plots for the **minimum expected shortfall** portfolio. The panels show annual performance measures for the DCC and the DCC-MVNCT model (left) and differences thereof (right) based on realized returns covering the years 2003 to 2012. On the right side differences are generally of the form “DCC minus DCC-MVNCT”. On the left side the results for the equally-weighted portfolio are overlayed for illustration.

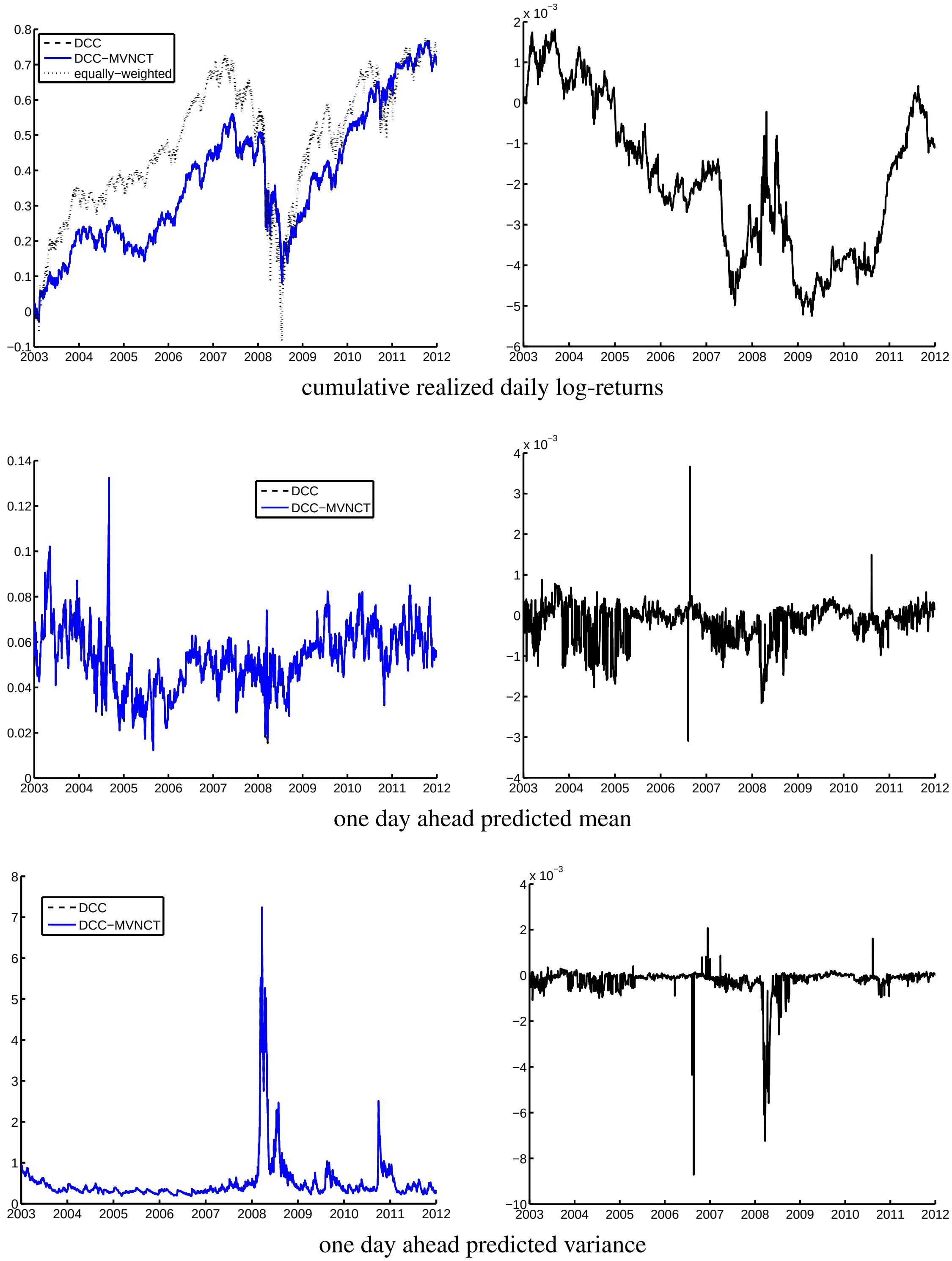
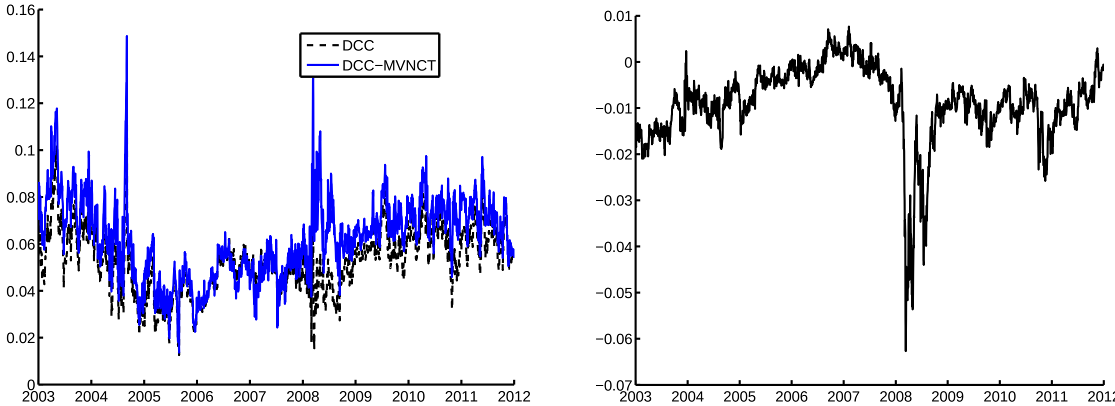
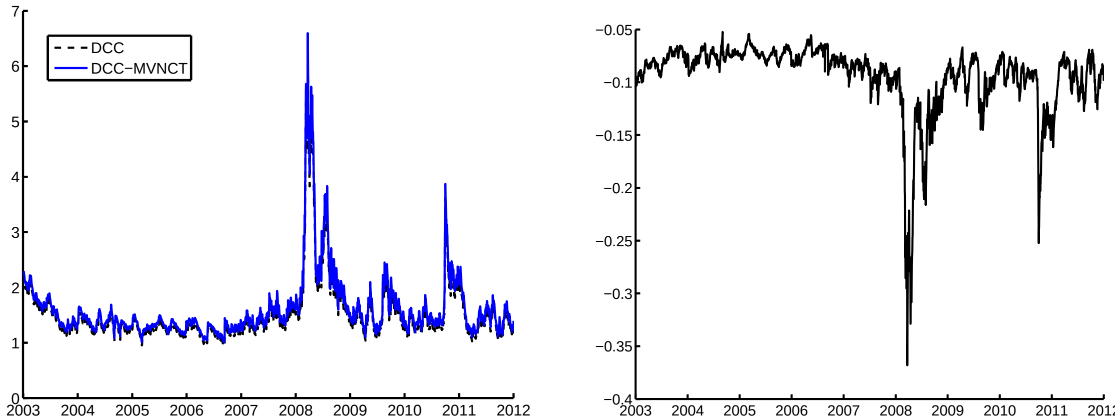


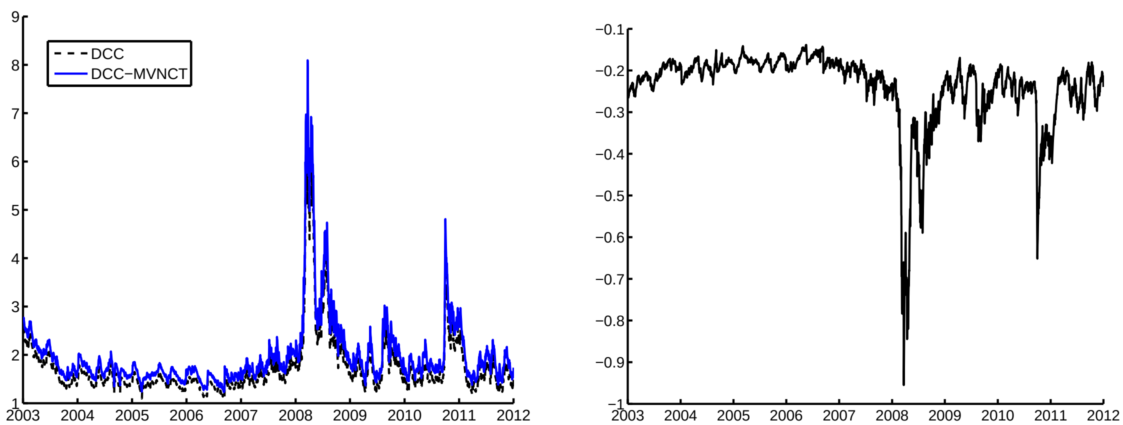
Figure 3.4: Cumulative returns and one-day-ahead predictions for mean and variance, covering the years 2003 to 2012. The upper left panel shows the evolution of cumulative returns over time for the **minimum expected shortfall** portfolio based on the DCC and the DCC-MVNCT model. For illustration the equally-weighted portfolio is included. As in Figure 3.3 differences between DCC and DCC-MVNCT are highlighted on the right side. The middle and lower panels depict the one-day-ahead predictions of mean and variance, respectively, for the min-ES portfolio.



one day ahead predicted median



one day ahead predicted 1% VaR



one day ahead predicted 1% ES

Figure 3.5: One-day-ahead predictions for median, 1% VaR and 1% ES, covering the years 2003 to 2012. The panels show results for the **minimum expected shortfall** portfolio based on the DCC and the DCC-MVNCT model. As in Figure 3.3 differences between DCC and DCC-MVNCT are presented on the right side. The predicted median corresponds to the 50% VaR.

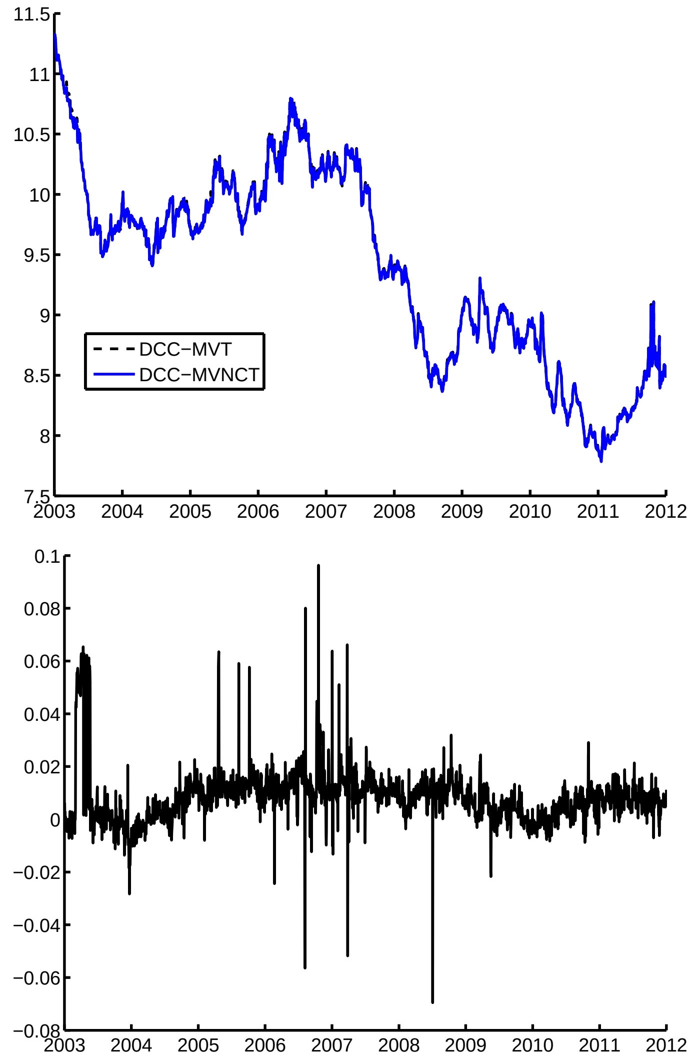


Figure 3.6: Evolution over time of the degrees of freedom of the DCC-MVT (ν_{T}) and the DCC-MVNCT (ν_{NCT}) model (upper panel), and differences, $\nu_{\mathrm{T}} - \nu_{\mathrm{NCT}}$, thereof (lower panel).

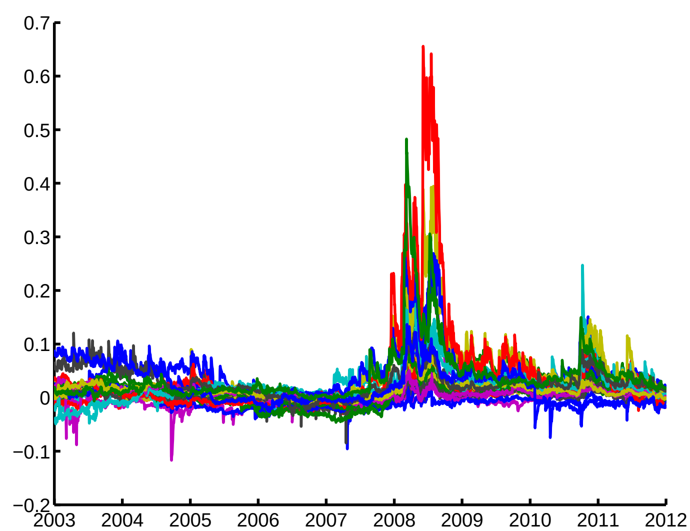


Figure 3.7: Evolution over time of the noncentrality parameters, γ, of the DCC-MVNCT model.

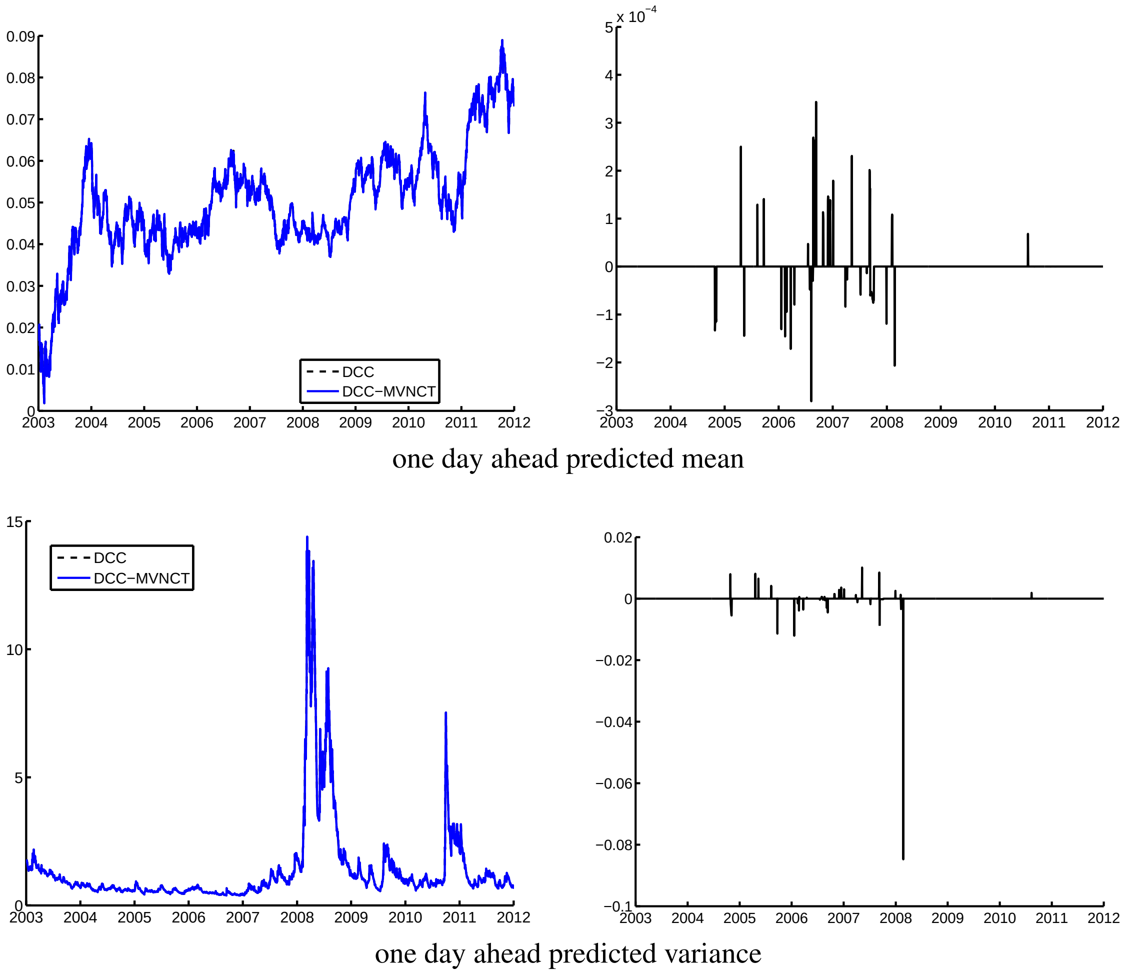
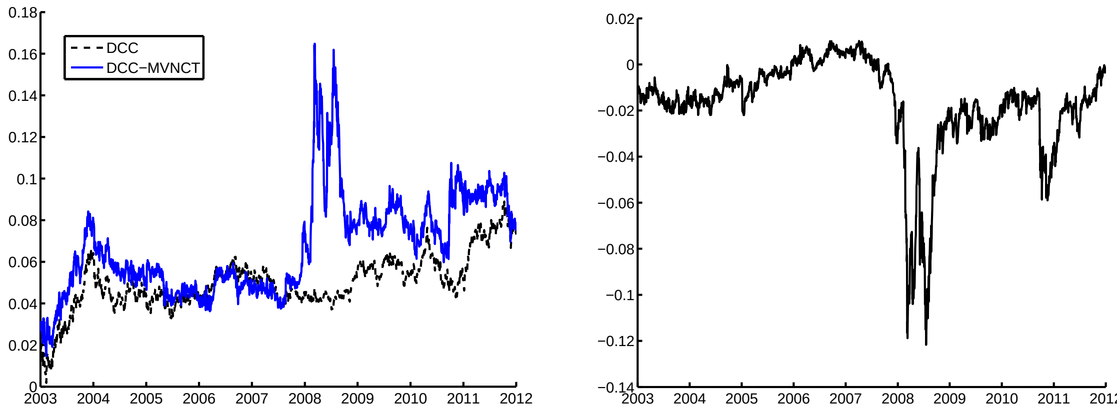
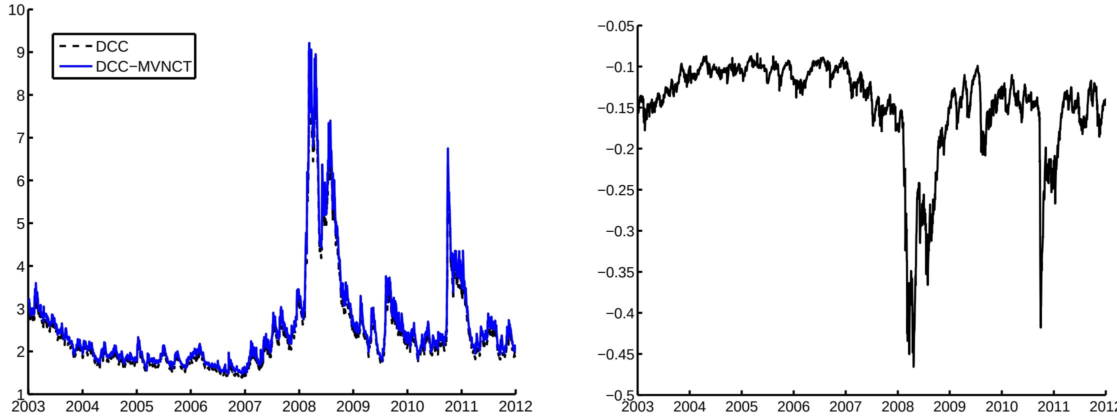


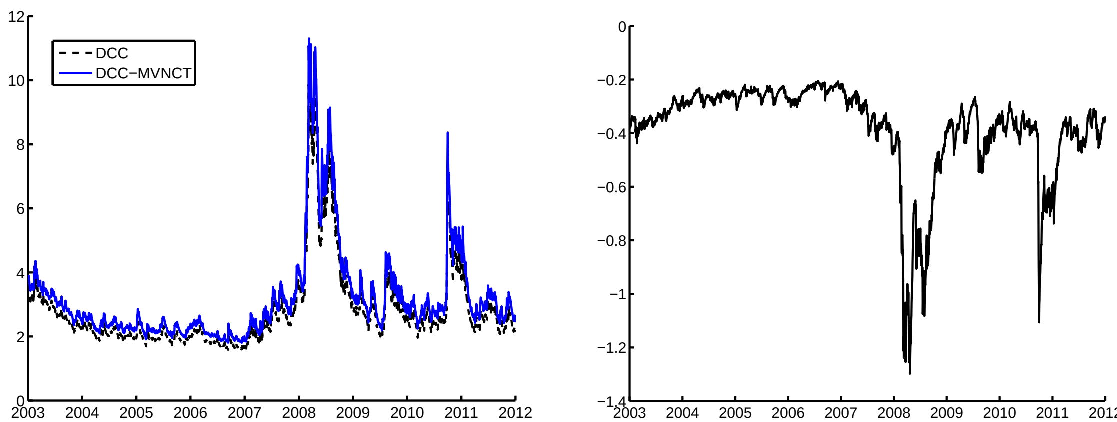
Figure 3.8: One-day-ahead predictions for mean and variance, covering the years 2003 to 2012. The panels show results based on the DCC and the DCC-MVNCT model for the **equally-weighted** portfolio. As in Figure 3.3 differences between DCC and DCC-MVNCT are given on the right side.



one day ahead predicted median



one day ahead predicted 1% VaR



one day ahead predicted 1% ES

Figure 3.9: One-day-ahead predictions for median, 1% VaR and 1% ES, covering the years 2003 to 2012. The panels show results based on the DCC and the DCC-MVNCT model for the **equally-weighted** portfolio. As in Figure 3.3 differences between DCC and DCC-MVNCT are emphasized on the right side. The predicted median corresponds to the 50% VaR.

Bibliography

Aas, K., Haff, I. H., and Dimakos, X. K. (2006). Risk Estimation Using the Multivariate Normal Inverse Gaussian Distribution. *Journal of Risk*, 8(2).

Alexander, C. and Lazar, E. (2006). Normal mixture GARCH$(1, 1)$: applications to exchange rate modelling. *Journal of Applied Econometrics*, 21(3):307–336.

Bonato, M. (2012). Modeling Fat Tails in Stock Returns: A Multivariate Stable-GARCH Approach. *Computational Statistics*, 27(3):499–521.

Broda, S. A. (2013). Tail Probabilities and Partial Moments for Quadratic Forms in Multivariate Generalized Hyperbolic Random Vectors. Discussion Paper 13-001/III, Tinbergen Institute, Amsterdam.

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013). Stable mixture GARCH models. *Journal of Econometrics*, 172(2):292–306.

Broda, S. A. and Paolella, M. S. (2007). Saddlepoint Approximations for the Doubly Noncentral t Distribution. *Computational Statistics & Data Analysis*, 51:2907–2918.

Broda, S. A. and Paolella, M. S. (2010). Saddlepoint Approximation of Expected Shortfall for Transformed Means. *UvA Econometrics Discussion Paper 2010/08. University of Amsterdam.*

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Daniels, H. E. and Young, G. A. (1991). Saddlepoint Approximation for the Studentized Mean, with an Application to the Bootstrap. *Biometrika*, 78:169–179.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13(3):253–265.

Embrechts, P., McNeil, A., and Straumann, D. (1999). Correlation And Dependence In Risk Management: Properties And Pitfalls. In *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press.

Engle, R. F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics*, 20:339–350.

Geweke, J. and Amisano, G. (2009). Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. *Center for Quantitative Economic Research, Working Paper Series, CQER Working Paper 09-04*.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.

Jondeau, E. (2010). Asymmetry in Tail Dependence of Equity Portfolios. *National Centre of Competence in Research, Financial Valuation and Risk Management, Working Paper No. 658*, (No. 658).

Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press.

Kshirsagar, A. M. (1961). Some Extensions of the Multivariate Generalization t distribution and the Multivariate Generalization of the Distribution of the Regression Coefficient. *Proceedings of the Cambridge Philosophical Society*, 57:80–85.

Ku, Y.-H. H. (2008). Student-t Distribution Based VAR-MGARCH: An Application of the DCC Model on International Portfolio Risk Management. *Applied Economics*, 40:1685–1697.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Martin, R. D., Rachev, S. T., and Siboulet, F. (2003). Phi-alpha Optimal Portfolios and Extreme Risk Management. *Willmott Magazine of Finance*.

Mencía, J. and Sentana, E. (2009). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *Journal of Econometrics*, 153:105–121.

Neyman, J. (1937). Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift*, 20:149–199.

Paolella, M. S. (2010). ALRIGHT: Asymmetric Large-Scale (I)GARCH with Hetero-Tails. *Swiss Finance Institute Research Paper Series N10 27*. Available at SSRN.

Paolella, M. S. and Polak, P. (2013). COMFORT: A Common Market Factor Non-Gaussian Returns Model. Revise and resubmit for the Journal of Econometrics.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of Conditional Value at Risk. *Journal of Risk*, 2:21–41.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23:470–472.

Sharpe, W. F. (1966). Mutual Fund Performance. *Journal of Business*, 39(1):119–138. Part 2: Supplement on Security Prices.

Sortino, F. A. and Price, L. N. (1994). Performance Measurement in a Downside Risk Framework. *Journal of Investing*.

Temme, N. M. (1982). The Uniform Asymptotic Expansion of a Class of Integrals Related to Cumulative Distribution Functions. *SIAM Journal on Mathematical Analysis*, 13:239–253.

Part III

Curriculum Vitae

Lars Jochen Krause

born September 26, 1978, in Cuxhaven, Germany

Educational Background

2014 – present	Post-doctoral Research Fellow, Empirical Finance, University of Zurich, Switzerland
2009 – 2014	Ph.D. candidate, Banking and Finance, University of Zurich, Switzerland
2008	B.A./M.A., Computer Science, University of Kiel, Germany
2007 – 2009	Research Assistant, Empirical Finance, University of Zurich, Switzerland
2006 – 2007	Exchange student, Bioinformatics, Nanyang University of Technology, Singapore
2005	B.A./M.A., Economics, University of Kiel, Germany
2001 – 2008	Studies in Computer Science, University of Kiel, Germany
2001 – 2005	Studies in Economics, University of Kiel, Germany
1999 – 2001	Studies in Economics, University of Göttingen, Germany
1998	University-entrance diploma, SZ Geschwister Scholl, Bremerhaven, Germany

Work Experience

2014 – present	Research Associate, Chair of Empirical Finance, Prof. Marc S. Paolella Department of Banking and Finance, University of Zurich, Switzerland
2007 – 2014	Research Assistant, Chair of Empirical Finance, Prof. Marc S. Paolella Department of Banking and Finance, University of Zurich, Switzerland
2007	Fair Assistant, Salomon HitBurger Gmbh, Großostheim, Germany
2005 – 2007	Teaching Assistant, Institute of Cognitive Systems, Prof. Gerald Sommer Department of Computer Science, University of Kiel, Germany
2005 – 2007	Temporary Personnel, Second German Television, Mainz, Germany
2003 – 2007	Research Assistant, Applied Statistics and Financial Econometrics, Prof. Stefan Mittnik Institute for Statistics and Econometrics, University of Kiel, Germany
2000 – 2001	Semester Assistant, Chair of Retailing, Prof. Bartho Treis Institute of Marketing and Retailing, University of Göttingen, Germany
1998 – 1999	Civilian Service, Rose 12, Addiction Service, Oldenburg, Germany